DOCUMENT RESUME

ED 110 362 95 SO 008 438

AUTHOR Golledge, R. G.; Rushton, Gerard
TITLE Multidimensional Scaling: Review and Geographical Applications, Technical Paper No. 10.
INSTITUTION Association of American Geographers, Washington, D.C. Commission on College Geography.
SPONS AGENCY National Science Foundation, Washington, D.C.
REPORT NO TP-10
PUB DATE 72
NOTE 89p.

EDRS PRICE MF-$0.76 HC-$4.43 PLUS POSTAGE
DESCRIPTORS Algorithms; College Instruction; Evaluation Methods; Geographic Distribution; Geographic Location; *Geography; Higher Education; Illustrations; Literature Reviews; *Maps; Mathematical Applications; Mathematical Concepts; *Methods; *Multidimensional Scaling; Social Sciences

ABSTRACT.

The purpose of this monograph is to show that sufficient achievements in scaling applications have been made to justify serious study of scaling methodologies, particularly multidimensional scaling (MDS) as a tool for geographers. To be useful research, it was felt that the common methodological and technical problems that specialized researchers share with other scholars should be indicated by review of the applications, and that an adequate statement on the mathematics and heuristics of scaling algorithms is necessary. As a review of applications, subroutines in scaling programs are "dissected" in order to understand how certain critical parameters are defined and used. This research work is presented in three parts relating to 1) basic fundamentals of scaling, data requirements, and algorithm constructions and problems; 2) two step-by-step examples of the non-metric section of a multidimensional scaling algorithm; and 3) a review of geographical applications of the approach in a variety of problem areas. The position of this paper is that MDS provides a useful and constructive methodology for examining the problems of preference and choice for researchers in geography. In conclusion, some problems of using MDS are mentioned and its potential uses in geography given. (Author/ND)

GEOGRAPHIC TECHNICAL PAPER SERIES

MULTIDIMENSIONAL SCALING: REVIEW AND GEOGRAPHICAL APPLICATIONS

R. G. Golledge
Gerard Rushton

COMMISSION ON COLLEGE GEOGRAPHY
TECHNICAL PAPER NO. 10

ASSOCIATION OF AMERICAN GEOGRAPHERS
WASHINGTON, D.C.

Supported by a grant from the National Science Foundation

ASSOCIATION OF AMERICAN GEOGRAPHERS
Commission on College Geography Publications

General Series

*No. 1–Geography in Undergraduate Liberal Education, 1965
No. 2–A Basic Geographical Library–A Selected and Annotated Book List for American Colleges, 1966
*No. 3–Geographic Manpower–A Report on Manpower in American Geography, 1966
No. 4–New Approaches in Introductory College Geography Courses, 1967
*No. 5–Introductory Geography–Viewpoints and Themes, 1967
No. 6–Undergraduate Major Programs in American Geography, 1968
No. 7–A Survey Course: The Energy and Mass Budget at the Surface of the Earth, 1968
No. 8–A Systems Analytic Approach to Economic Geography, 1968
No. 9–A Geographical Bibliography for American College Libraries, 1970
No. 10–Geography in the Two-Year Colleges, 1970

*Out of print

Resource Paper Series

No. 1–Theories of Urban Location, 1968
No. 2–Air Pollution, 1968
No. 3–Perspectives on Geomorphic Processes, 1969
No. 4–Spatial Diffusion, 1969
No. 5–Perception of Environment, 1969
No. 6–Social Processes in the City: Race and Urban Residential Choice, 1969
No. 7–The Spatial Expression of Urban Growth, 1969
No. 8–The Political Organization of Space, 1971
No. 9–An Introduction to Spatial Allocation Analysis, 1971
No. 10–Man and Nature, 1971
No. 11–Tropospheric Waves, Jet Streams, and United States Weather Patterns, 1971
No. 12–The Spatial Structure of Administrative Systems, 1972
No. 13–Residential Mobility in the City, 1972
No. 14–The Periglacial Environment, Permafrost, and Man, 1972
No. 15–Conservation, Equilibrium, and Feedback Applied to Atmospheric and Fluvial Processes, 1972
No. 16–Metropolitan Neighborhoods: Participation and Conflict Over Change, 1972
No. 17–Computer Cartography, 1972
No. 18–Society, The City and The Space-Economy of Urbanism, 1972
No. 19–Thematic Cartography, 1972

Continued on inside back cover

MULTIDIMENSIONAL SCALING: REVIEW AND GEOGRAPHICAL APPLICATIONS

by

R. G. Golledge
Ohio State University

and

Gerard Rushton
University of Iowa

ASSOCIATION OF AMERICAN GEOGRAPHERS
Commission on College Geography
Washington, D.C. 20009

TECHNICAL PAPER NO. 10

Library of Congress Catalog Card Number 72-90875

Supported by a grant from the National Science Foundation

FOREWORD

The Technical Papers are explanatory manuals for the use of both instructors and students. They are expository presentations of available information on each subject designed to encourage innovation in teaching methods and materials. These Technical Papers are developed, printed, and distributed by the Commission on College Geography under the auspices of the Association of American Geographers with National Science Foundation support. The ideas presented in these papers do not necessarily imply endorsement by the AAG. Single copies are mailed free of charge to all AAG members.

John F. Lounsbury
Project Director
Commission on College Geography

Paul W. English
Chairman, Panel on Resource
and Technical Papers

Panel Members

Ian Burton, University of Toronto
Paul W. English, University of Texas
Leslie J. King, McMaster University
Harold M. Rose, University of Wisconsin, Milwaukee
Robert H. T. Smith, Queen's University
Edward W. Soja, Northwestern University
Philip L. Wagner, Simon Fraser University
Julian Wolpert, University of Pennsylvania

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

An invitation from the Commission on Quantitative Methods, International Geographical Union, in November 1969, "to prepare for 'Multivariate Scaling' a detailed discussion of the technique and the reason for its use, ending with a step-by-step worked example using real data," for presentation at a meeting of the I.G.U. Quantitative Methods Commission in Poznan in September 1970, appeared to us to be an incentive to weave together the patchy knowledge of scaling algorithms that we had found scattered throughout many periodicals and manuscripts.

For geographers who look askance at the necessity of becoming acquainted with yet another statistical tool, we can only express our conviction that the importance of the scaling problem has been too little recognized in geography and that, insofar as a major area of geographic interest is with choice data, scaling techniques provide the tools for analyzing such data. As earlier styles in geographic research emphasizing structural descriptions give way to research styles that search for process laws first and then explore their spatial implications, so techniques that are suited to the analysis of choice and the preference functions underlying choice, must be adopted. In comparison with the parametric statistical techniques more commonly used by geographers, non-metric scaling techniques allow the researcher to be more flexible in searching for functional forms and in designing schemes for assembling basic data. To take advantage of this flexibility is both a problem and a challenge. Our hope is that this monograph will indicate to the skeptic that sufficient achievements in scaling applications have already been made to justify the serious study of the still developing area of scaling methodologies.

As with any basic analytical technique, applications range through every subfield of geography. Consequently we felt that a review of these applications might be useful in indicating to researchers immersed in their specialized area, the common problems of a methodological and technical character that they share with a far wider community of scholars. In preparing the review, however, we felt that an adequate statement on the mathematics and heuristics of scaling algorithms was first necessary. Ideally, we would have preferred to make a citation to a review that already existed. Search as we did, we were not able to locate a review which served our purposes, though we saw some references to reviews "in preparation." The excellent treatments of Torgerson [57] and Coombs [7] largely predated the major developments of Shepard [50, 51] and of Kruskal [28, 29, 30]. These, in turn, spawned further theoretical and empirical works that we soon found were not described adequately in print even though their understanding was critical for those researchers who depended on scaling in their substantive work. We found that information on this subject was being transmitted through numerous working papers, discussion papers, unpublished dissertations, newsletters, and computer program printouts. In our own research we found it necessary to "dissect" subroutines in scaling programs in order to understand how certain critical parameters were being defined and used. Thus the work is presented in three parts relating to basic fundamentals of scaling, data requirements, and algorithm constructions and problems, two step-by-step worked examples of the non-metric section of a multi-

dimensional scaling algorithm, and a review of geographical applications of the approach in a variety of problem areas.

The willingness of some of the most prominent researchers in multidimensional scaling to send us unpublished manuscripts, reprints, computer printouts, and test data, and, in the case of Professor Lingoes, to provide us with a most helpful critique of the I.G.U. paper, was a source of inspiration to us to complete the work. In addition, several of our students helped us by working on some of the programming problems and by running experimental data through the various algorithms. Although we had intended to experiment further with different algorithms and to report results in this paper, we found that Lingoes and Roskam [33] and Young and Applebaum [62] had admirably designed experiments and reported on this question.

We thank Professor James Lingoes of the University of Michigan and Professor Forrest Young of the University of North Carolina for the materials they sent us and for their most interesting communications. We thank also Professor Waldo Tobler of the University of Michigan for his interest in the work, for generously allowing us to use his trilateration example in Section I, and for showing us the variety of map transformations applications of scaling. At the University of Iowa, Mr. John F. Hultquist and Mr. Stanley R. Lieber have helped us by testing and using the TORSCA scaling algorithm; at McGill University, Dr. Gordon Ewing helped us by describing and commenting on his use of the Guttman-Lingoes SSA-I algorithm, at Ohio State University Dr. Ronald Briggs and Dr. Donald Demko (now at the University of Texas and Queen's University, respectively) experimented with unfolding techniques and with the Kruskal MDSCAL series of programs; Professor L. Neidell formerly of the Department of Marketing, SUNY at Buffalo, drew our attention to several interesting spatial applications of scaling methods outside the field of geography and provided tapes of the entire Guttman-Lingoes series, Professor Paul Isaacs of the Psychology Department, The Ohio State University, gave critical technical advice on the interpretation of various algorithms. We also acknowledge that the Computer Institute for Social Science Research, Michigan State University, provided the initial stimulus and the programming assistance that led to our applying multidimensional scaling techniques in our own work. Finally, we acknowledge our gratitude to Dr. T. Czyz, Dr. Z. Chojnicki, and other Polish geographers who created such a convivial atmosphere for discussion at the I.G.U. meeting in Poznan, September 1970.

Acknowledgements are also due to the following people and organizations for permission to use diagrams and maps. Professor W. Tobler, University of Michigan, Northwestern University Press and Department of Geography; Association of American Geographers, Canadian Association of Geographers, Dr. Paul Schwind, University of Hawaii, *Geographical Analysis,* and Professor Peter Gould, Pennsylvania State University.

Additional support was granted to Rushton by the Graduate College, University of Iowa, and the United States Army Research Center, and to Golledge by the Departments of Geography, University of Sydney, Australia, and The Ohio State University.

I. AN INTRODUCTION TO SCALING AND SCALING ALGORITHMS

A. Scaling

The fundamental idea of scaling is to produce a range of scores that have meaning either with respect to each other's values or to some arbitrary or absolute value; set or accepted by the scale. A scale generally consists of a system of numbers related by correspondence rules which enable meaning to be attached to the objects possessing them. For example, a number system is a scale which can be nominal, ordinal, interval, or ratio in nature [27, pp. 9-12]. The explication of any number system involves itemizing the correspondence rules which give meaning to each number in the system [41]. Thus, we can envisage that potentially there is an infinity of ways to scale data, generally, however, the scaling problem reduces to one of devising rules for the measurement of a construct or phenomenon such that the resulting measurements provide an easily interpretable and admissible transformation into numerical form of the phenomenon being scaled.

The advantages of scaling are similar to those which derive from the measurement of properties in general:

1. scaling makes it possible to differentiate among instances that may be lumped together in a given class of things (e.g., degrees of "warmth");
2. scaling can show relative position rather than just difference; and
3. scaling allows systematic manipulation of the scaled items in conformance with concepts and theories of logic and mathematics.

Scaling methods are generally subdivided into two classes - unidimensional and multidimensional. Unidimensional scales are those which measure variation with respect to one attribute (e.g., color hue, population size, neighborliness, per capita income, residential status, social rank, degree of urbanization, and so on). For example, Berry and Ginsberg [4] constructed a series of unidimensional scales measuring a variety of urban, economic, and societal characteristics of various countries of the world and used those scales in a multivariate analysis of levels of economic development. Similarly, the urban rank size rule and the indices of primacy developed by Linsky [35] and Mehta [40] are unidimensional scales based on population-size which provide a range of scale values. However, despite the widespread use of such scales in geography, we must be aware of the problem that systematic variation of the scaled phenomena with respect to more than the assumed number of attributes may be undetected and may lead to difficulties in interpreting the results, or to a rejection of experimental results. Examples might include divergences between actual migrant moves and scales expressing the "desirability" of places of settlement, or variations in the attracting power of places resulting from variations in the number of functions rather than just population size [3].

Multidimensional scaling aims at developing procedures which will assign sets of numbers to various quantities of attributes such that the numbers directly reflect variations in the quantities of the attributes among the phenomena being scaled.

For example, distinctions may be called for between objects which vary with respect to color, location, shape, size, and so on. Such objects form a multidimensional series, and the scaling procedure attempts to identify a number of relevant dimensions of these objects. To do this, data are collected to permit the detection of systematic variation with respect to more than an assumed small number of attributes. For example, if scaling is used as an evaluative device, a subject is generally instructed to consider the similarity or dissimilarity of stimulus objects (considering all relevant attributes) rather than being asked to rate objects according to the magnitude of some specific attribute. Specific examples of these procedures are discussed later in the paper.

Recognition that any object may have a number of attributes, and that different attributes may be used by different individuals in their attempts to scale the objects in some way, led to the conclusion that any given object could be regarded as existing in an n-dimensional space, where n represents the number of perceived or actual attributes. The quantity of each attribute belonging to an object can then be interpreted as a geometrical coordinate which, when used in conjunction with other quantities (coordinates), determines the location of each object in the n-dimensional space. The significance of this is that if individual objects can be thus located, then interpoint distances can be calculated and objective statements can be derived concerning the distances separating various objects.

There are some important points to consider about this geometrical interpretation of the multidimensional scaling process. First, consider the case where objects are located in an n-dimensional real Euclidean space. Here each number associated with an attribute gives the projection of the object on one of the coordinate axes of the space, in other words, it allows us to determine the distance of the object from an origin along a given axis. The distance between two points located in this space is given as follows:

$$d_{jk} = \left[\sum_{r=1}^{n} (p_{rj} - p_{rk})^2 \right]^{1/2} \qquad (1)$$

where j and k are the two points,
r is an index of the axes,
n is the number of orthogonal axes,
and p_{rj}, p_{rk} are the projections of the points on axis r.

What is particularly important is that given the distances between all pairs of points in the space, the projections of the points on any arbitrary set of orthogonal axes in the space can be determined. In other words, given interpoint distances, we can recover the number of dimensions in which the points exist. For any set of interpoint distances there will be a space of minimum dimensionality in which a satisfactorily large number of the interpoint distances maintain their relationships one to another. One of the aims of multidimensional scaling is to identify this space of minimum dimensionality and to interpret each dimension in terms of stimulus attributes.

The second critical feature of multidimensional scaling is that it is not necessary

to have metric information on the interpoint distances as input. Because subjects are not asked to make decisions with respect to a given number of attributes, they are free to choose any number of attributes they desire in order to make distinctions between objects. Thus, instead of imagining that each stimulus object has a location in a real Euclidean space, the subject locates each object in a "psychological space" which may exist in a quite different form of geometrical space. In addition, subjects locate points in the space merely in terms of their being "nearer," "greater than," "more similar," "more preferred," (and so on) to any given object than they are to others. The aim of multidimensional scaling is to take the data collected with respect to stimulus objects and to recover from these data a spatial configuration of points in an identifiable space of minimum dimensionality.

B. Data Requirements and Collection

It was suggested above that multidimensional scaling procedures are versatile in that they can take as input either metric or non-metric data. Since discussions of the various types of metric data can be found in any basic mathematics or statistics book, emphasis will be placed here on using non-metric information.

Coombs [7] argues that there are four basic kinds of behavioral data—preferential choice, single-stimulus, stimulus-comparison, and similarities data.

Assume that we have a sample of individuals and a collection of phenomena and that the individuals are asked to state their preferences for the phenomena. The instructions may be of the following types:

1. Choose one out of a set of n phenomena.
2. Choose k out of a set of n phenomena.
3. Choose one of a series of subsets of the n phenomena.
4. Choose one out of every possible pair of the phenomena.

When sample members perform one or another of the above tasks they state their preferences for the phenomena chosen, and the data collected are called *preferential choice data.* For this type of data, we assume that the set of stimuli and the individuals revealing their preferences can be mapped together into a joint space. This joint space is a psychological space, and both stimuli and individuals are mapped as points in such a way that the relations among the different points in the space reflect the observed preference orderings of the individuals. It can be hypothesized that each of the individuals who has been asked to state preferences among the stimuli will have different ideals as to what an appropriate combination of stimulus attributes should be. Thus, if a particular stimulus has more or less of a particular attribute, individual preferences might reveal it to be more or less desired. Through a sequence of preferential choice statements, an individual reveals where each stimulus point lies with respect to his ideal point. We can imagine that the same basic dimensions are used by most individuals in order to make judgments about the stimuli but that each individual may prefer different quantities of each attribute. Thus, the attributes that are perceived in any stimulus will be mapped into the space in proportion to whether or not an individual desires that particular

attribute, in other words, in relation to where his ideal point lies with respect to any dimension by which a given attribute is represented. Given a sample of individuals and a sample of stimuli, the stimuli may be mapped into the same space as the individuals. The problem then is to find how the individuals and the stimuli can be mapped into a space of minimum dimensionality to reflect the preference orderings of each of the individuals.

This particular problem becomes complicated somewhat when we realize that when choosing between pairs of phenomena, for example, subjects will not always reveal the same preferential ordering. Indeed, in an experiment where a large number of preferential choices are made, each individual may make intransitive choices, that is, he may prefer A to B, B to C, but C to A. If there are no such intransitivities in the preference ordering, then stimuli can be rank ordered for each individual from most to least preferred and a unidimensional scale for that individual can be compiled. While the problem still remains to fit together all the individual unidimensional scales, it becomes far more complicated if any given individual has intransitivities in his preference orderings. Intransitive preference rankings (or preference orderings) of objects cannot be arranged in simple rank order. Intransitive preference orderings may occur when a stimulus object with multiple attributes evokes in a respondent an ordering based on a particular dominant attribute for one comparison, and then evokes a rank ordering on the basis of a different attribute for a second comparison. For example, candies may be considered to have the attributes of brittleness, chewiness, quality of chocolate, or type of filling. Given any two candies, an individual may compare them from the point of view of brittleness but when comparing either of them with another, an alternative attribute such as type of filling may be used as the distinguishing criterion.

Coombs' idea with respect to this preferential choice problem is to *unfold* the space in which preferential choices are being made such that both individuals and objects are mapped as points in a joint space and the mutual relations among the points reflect the observed preference orderings of the various individuals [7, pp. 80-192]. The unfolding of choices is not our major concern in this paper, but it does provide one basis for the multidimensional scaling algorithm developed by Coombs, which is referred to later in the paper.

The second type of data that Coombs defines is *single stimulus data*. Here we assume that our sample individuals are presented with a set of homogeneous stimuli (i.e., stimuli from a single population such as political candidates, supermarkets, etc.), only now the individual is asked to make a judgment about each stimulus in turn. For example, one may be asked whether or not he would vote for a certain politician. In this case, if we again regard each individual as having an ideal point then we can imagine that each stimulus when presented to the individual is said to lie either within the "neighborhood" of his ideal point or outside this hypothetical neighborhood. This would generate his yes-no response. These data are sometimes described in abstract terms as "proximity relations" and are somewhat different from the order relations expressed in preferential choices.

A third type of data arises when we ask individuals to determine an order

relation on pairs of points from the same data set: these are called *stimulus comparison data* and the method of collection is generally the method of paired comparisons. Consider the candy example mentioned previously. Assume that we suggest every possible pair of candies to our subjects and that we ask each individual to judge between the two candies on the basis of some actual or perceived characteristic. In the simplest of all experiments we would simply choose a single attribute such as brittleness, for example, and ask for comparison between the candies on the basis of this attribute. Judgments would therefore be presumed to reflect differences among stimuli and not among individuals. If, however, we find that the individuals do not agree with respect to their interpretation of the attributes, then a frequent technique is to subdivide the population such that the groups created are homogeneous with respect to their perceptions of the attributes. Note that in comparison to the two previous procedures the individual need not be interpreted as a point in the same space as the stimulus. Rather, the comparisons between the stimuli enable us to locate them in the space of some dimensionality such that the distance between each pair of points in the space is interpreted as an indication of the similarity or dissimilarity of the objects. Short distances indicate high degrees of similarity and large distances denote dissimilarity.

The method of collecting paired comparison data can differ somewhat from experiment to experiment. In stimulus comparison experiments, frequent use is made of paired comparison, triadic comparisons, and comparisons of pairs of dyads [7, pp. 3-59, 444-462] The typical paired comparison experiment involves subjects being given all possible pairs of objects and then being asked for some type of comparative judgment related to the pairs. If it is assumed that self-similarities (i.e., comparing the object to itself) are ignored and complementary comparisons are equivalent (i.e., A-B is assumed the same as B-A), then there are $\frac{n(n-1)}{2}$ pairs of objects for which some type of comparative judgment is requested. The ultimate aim is to obtain the ordering of the paired objects upon the basis of some psychological continuum. The procedure is based on one of the fundamental principles of the Law of Comparative Judgment. This states that any given stimulus has associated with it a most frequently aroused or modal discriminal process on some continuum. It is accepted that any subject may choose different attributes of the stimulus object when comparing the object with others, but it also assumes that the discriminal process (or reaction of the subject to the stimulus) is distributed normally around a mode which can be called the *scale value* of the object. Thus, any two objects may differ with respect to their scale values. The purpose of the paired comparison experiment then is to allow the scale values for any two stimulus objects to be compared so that statements can be made about the degree of similarity or difference that they evoke in any respondent. By finding the frequency with which stimulus *i* exceeds, equals, or is rated less than stimulus *j* we obtain judgment as to the relative magnitudes of their respective scale values.

In stimulus comparison experiments, the stimulus attributes upon which judgments are made are clarified for the individual. If we permit the subject freedom of

choice in comparing stimuli, this leads to the development of a fourth kind of data called *similarity data*. With this type of data the individual is presumed to perceive each stimulus as a union of sets of attributes. Again, it is assumed that the stimulus is able to be represented by a point in space, the coordinates of this point correspond to the projections of the stimulus on the various dimensions which the individual chooses as being relevant. The object of making this type of comparison is to attempt to determine the minimum number of dimensions which are used by individuals when comparing objects. For this type of data, comparisons are made usually by asking subjects if one pair of stimuli are more nearly alike than another pair. In other words, we attempt to find out whether the distance separating one pair of stimulus points is less than, equal to, or greater than the distance separating another pair of corresponding points. In a sense, an attempt is made to determine an order relation on the distances implicit between pairs of points where all the points are from a single homogeneous set. Multidimensional scaling models are commonly used with similarities data in an attempt to construct some type of stimulus space from measures of the interpoint distances between phenomena.

Of the various types of data available, geographers have experimented to some extent with each one, although the use of paired comparison procedures for collecting data is becoming somewhat more popular. Examples of such experiments include asking consumers to select one of a pair of shopping centers, towns to visit in order to purchase given goods, or towns which would be selected for purposes of migration. The frequency with which any given pair member is chosen over others is then recorded [5,45]. For example, Rushton interpreted the movement of farmers to towns in Iowa as the outcome of a choice process which could be inferred to be a paired comparison type procedure. Visits to each place were transformed into dissimilarities measures. first, by recording the number of times town type i was chosen (for a specified shopping trip) over town type j when both i and j were present in a feasible area; second, by regarding a proportion of 0.5 as being the maximum perceived similarity, and then by finding the difference between the derived proportions and this maximum perceived similarity:

$$d_{ij} = \left| p_{ipi} - 0.5 \right| \qquad (2)$$

where d_{ij} is a measure of dissimilarity. Here a small value for d_{ij} represented small dissimilarity between towns, and a large value indicated considerable dissimilarity.

Most multidimensional scaling analyses use only the upper half of a paired comparison matrix. Thus only $\frac{n(n-1)}{2}$ comparisons are used as input. Most computing algorithms, however, allow optional inclusions for diagonal and lower half matrices and they can be used either with complete or incomplete data.

C. Approaches to Non-Metric Multidimensional Scaling (MDS)

It was previously suggested that implicit in every collection of proximity

measures (such as measures of similarity and dissimilarity) is a kind of spatial structure. The basic problem of MDS is to uncover this structure. While it is generally agreed that greater degrees of similarity infer closer distances (and vice versa), the former are only implied distances and may not easily be transformed into metric form. However, in Shepard's terms: "If some monotonic transformation of the proximity measures could be found that would convert these implicit distances into explicit distances, then we should be in a position to recover the spatial structure contained only latently in the original data" [50, p. 127].

The various approaches developed by authors such as Coombs, Torgerson, Shepard, Kruskal, McGee, Guttman, Lingoes, and Young represent attempts to recover the latent spatial structures contained in proximity-type data [7, 57, 51, 29, 38, 39, 18, 32, 63].

The relative advantages of the non-metric approach in searching for latent spatial structure have been summarized by Lingoes and Guttman:

> One of the chief benefits to be derived from constraining the solution non-metrically, is, of course, that in general a smaller space is required to reflect order than to reflect metric. Of greater importance, however, the dimensions themselves may well aid our understanding of the underlying interdependencies free of the attenuation that can result from non-linear relationships. Furthermore, when some lawful structure or pattern is present in the data, e.g., a simplex, a circumplex, or a radex, a non-metric analysis will reveal the configuration whereas a metric approach will obscure the lawfulness [32, p. 487].

1. **The basic elements of non-metric MDS algorithms.** Although there are differences among the algorithms currently in use, there are also broad similarities in terms of their construction. Features common to the majority of the techniques include:

a) an initial set of input data, frequently generated by a paired comparison experiment, within which is contained a latent spatial structure (such as dissimilarities data). These input data can be prepared in random vector mode, or in the form of a symmetric matrix, a rectangular matrix, or a triangular matrix.

b) an initial configuration of interpoint distances which is manipulated on successive iterations in an attempt to define a monotone relationship between the configuration and the original data.

c) a computing algorithm (a non-metric scaling method) which incorporates the strategy for achieving convergence of the data and the configuration.

d) a loss function (or "goodness-of-fit") function which is used to guide and/or terminate the iterative procedures.

e) subroutines for handling missing data and tied data, and for determining step size motions within each generated configuration.

f) techniques for estimating the configuration deformation as the number of dimensions in which the configuration is plotted is changed.

A generalized format for a non-metric MDS analysis of complete data has been provided (as follows) by Lingoes and Roskam [33, pp. 11-16].

Let Δ = a k-element array (or vector) of arbitrary indices of dissimilarity between all pairs of n-objects, $k = n(n-1)/2$

δ_{ij} = the general element of Δ, $(i = 1,2,\ldots,n-1;$
$j = i+1, i+2,\ldots,n)$

Let $\delta_{ij} = \delta_{ji}$, and ignore therefore symmetric elements and diagonal elements.
Define $\hat{D}$ as a k-element vector of real numbers with elements $\hat{d}_{ij} = f(P_{ij})$ such that, when $\delta_{ij} < \delta_{kl}$, either $\hat{d}_{ij} < \hat{d}_{kl}$, or $\hat{d}_{ij} \leqslant \hat{d}_{kl}$.

Note that $\hat{d}_{ij} < \hat{d}_{kl}$ implies semi-strong monotonicity when some Δ-elements are tied and strong monotonicity with no ties, and $\hat{d}_{ij} \leqslant \hat{d}_{kl}$ implies weak monotonicity for no ties and semi-weak monotonicity for ties (the following section discusses monotonicity).
Assume $\hat{D} \rightarrow \Delta$ monotonically: now $\hat{D}$ is a monotonic transformation of the Δ-vector whose function is to weight the iterations for moving a configuration toward its goal and to form a basis for evaluation of goodness-of-fit at any iteration.
Define X an an $n \times m$ matrix of rectangular coordinates (for a given configuration) with m representing the number of dimensions.
Define D as a k-element vector of distances calculated from X between the n-points embedded in a given space according to the standard distance formula given earlier as equation (1). The general problem statement can then be formalized as follows. given Δ, an initial configuration X, a fixed n, and distances calculated between the n points, try to get D as close as possible to $\hat{D}$ (i.e., minimize some loss function), for then D will map into Δ.

The general procedure is to:

1) determine an initial set of coordinates (X) and select an appropriate dimensionality (m);
2) compute the Euclidean distances (D);
3) solve for $\hat{D}$, the predictions of the appropriate distances $\{$i.e., $d_{ij} = f(\delta_{ij})\}$;
4) compute the normalized loss function (e.g., STRESS);
5) if the loss function is small enough (or not changing "sufficiently" from one iteration to the next), terminate. Otherwise modify X and return to step 2.

At this stage, it is pertinent to discuss some of the basic elements of MDS approaches as a means of explaining some of the rules which govern algorithm construction. In particular, we plan to focus on the requirement of monotonicity, methods of deriving initial configurations, goodness-of-fit criteria, treatment of ties, and some brief comments on dimensionality. Summaries of a selection of approaches to MDS will then indicate some differences in the strategies used to recover configurations from non-metric data.

2. Montonicity requirements. The essence of MDS algorithms is the requirement of maintaining a monotonic relation between the original dissimilarities data and the

distances which are derived to represent them. While it may seem at first that the imposition of the monotonicity restraint is a weak one, Shepard has argued that.

> If non-metric constraints are imposed in sufficient number they begin to act like metric constraints. In the case of a purely ordinal scale the non-metric constraints are relatively few and, consequently, the points on the scale can be moved about quite extensively without violating the inequalities (i.e., without interchanging any two points). As these same points are forced to satisfy more and more inequalities on the interpoint distances as well, however, the spacing tightens up until any but very small perturbations of the points will actually violate one or more of the inequalities [51, p. 288].

Thus, if we impose a monotonic constraint on the relationship between a set of dissimilarity measures and a set of distance measures, we are in effect assuming that the rank order of dissimilarities is by itself enough to obtain a solution—i.e., to obtain the latent spatial structure.

Monotonicity requirements can be specified in a number of ways [18, pp. 480-484]. Let R be the number of distinct values among the off-diagonal elements of a matrix consisting of an initial set of distance ranking numbers; $R = n(n-1)$ if, and only if, there is complete information (i.e., no missing data) and no ties. Where each number is tied only with its transpose in a symmetric matrix $R \leqslant \frac{n(n-1)}{2}$; $R = \frac{n(n-1)}{2}$ where information is complete and the only ties are with the transpose.

Now, given a trial matrix of distances, let Q be the number of off-diagonal elements which are specified (i.e., not missing). Because of the symmetry of the data, $Q \leqslant \frac{n(n-1)}{2}$. This allows for the definition of three types of monotonic conditions based on the potential presence or absence of ties and/or missing data:

1) if $Q = R$, *strong* monotonicity is said to exist (i.e., no ties)
2) if $Q \geqslant R$, *semi-strong* monotonicity is said to exist,
3) if $Q \leqslant R$, *weak* monotonicity is said to exist.

In general, permitting the removal of ties enables a smaller space to be attained than otherwise and most algorithms adopt conditions of either $Q \geqslant R$ or $Q \leqslant R$.

The first objective of the monotonicity constraint is to ensure that the rank order of interpoint distances obtained from some configuration of points is identical with the rank order of the original dissimilarity measures (or inversely monotonic with similarities data). Thus points with the smallest dissimilarity measures should end up with the smallest distances between them in the final spatial structure. A perfect monotonic relation would involve exactly the same ranking for corresponding pairs of dissimilarities and configuration distances.

Let δ_{ij} be a measure of dissimilarity between n "types" of stimuli [28]. For a matrix of such dissimilarities the intent is to represent the n types as n points in t

dimensional space, wherein the interpoint distances (d_{ij}) correspond to the observed degrees of dissimilarity between the n types. Perfect correspondence would mean, for example, that if type i is more similar to type j than it is to type k then the corresponding interpoint distances would satisfy the same relationship for all i, j, k, that is, where $\delta_{ij} > \delta_{ik}$, $d_{ij} > d_{ik}$. In other words, if the locational types are shown on a scatter plot in which the ordinate is dissimilarity (δ) and the abscissa is distance (d), then as the points are traced one by one from bottom to top, the move is always to the right, never to the left. When this requirement is met, a monotone relationship between dissimilarity and distance has been found.

In order to obtain appropriate rank orders for the distance measures, a monotone regression is performed and in this way, by minimizing the sums of squared differences between the derived set of numbers and the dissimilarities, it is ensured that the set of numbers representing the distances are "as much like" the dissimilarity measures as possible. In essence a monotone regression between dissimilarity and distance measures requires that only the distance measures be moved at each iteration (since the dissimilarities are dimensionless numbers). Differences between MDS algorithms arise when different methods are used to define the transformed distances and to perform the movement of the points. The critical problem is to determine the direction and magnitude of any moves of the points that have to take place; the two most favored methods are the "method of steepest descent" (or gradient method) and the "method of correction matrices" [29, 18].

3. Treatment of ties. Kruskal provides two options for the treatment of ties. The first or primary approach treats ties as an indeterminate order relation which can arbitrarily be resolved in such a way that either STRESS or dimensionality can be decreased. The secondary approach regards ties as being evidence of an equivalence relation which should, as far as possible, be maintained—even if the result is to increase dimensionality or stress [33, pp. 36-37].

In the primary approach—the one which Kruskal originally preferred [20, p. 22]—when two dissimilarities were equal (i.e., $\delta_{ij} = \delta_{kl}$) it was argued that it was of no great concern whether d_{ij} or d_{kl} was the larger, or whether in fact they were equal or not. Thus if $d_{ij} \neq d_{kl}$ there was no pressure to downgrade the configuration, the inequality was not reflected in the stress value, and no constraints were placed on the estimates of distance ($\hat{d}_{ij}$ & $\hat{d}_{kl}$). The terms $(d_{ij} - \hat{d}_{ij})^2$ and $(d_{kl} - \hat{d}_{kl})^2$ were therefore permitted to be zero (subject to the influence of any other existing constraints). Whenever $d_{ij} < d_{kl}$, the (transformed) distance estimates had to follow the same monotone relation or could be equal (i.e., $\hat{d}_{ij} \leqslant \hat{d}_{kl}$).

The secondary approach on the other hand argued that if $\delta_{ij} = \delta_{kl}$ then either $d_{ij} = d_{kl}$ or the configuration should be downgraded until the equality holds. Thus estimates of the distances must be equal ($\hat{d}_{ij} = \hat{d}_{kl}$). If $d_{ij} \neq d_{kl}$ then $(\hat{d}_{ij} - \hat{d}_{ij})^2$ and $(d_{kl} - \hat{d}_{kl})^2$ are not zero and are reflected in the stress value. In other words, whenever $\delta_{ij} < \delta_{kl}$ then $\hat{d}_{ij} \leqslant \hat{d}_{kl}$, and whenever $\delta_{ij} = \delta_{kl}$ then $\hat{d}_{ij} = \hat{d}_{kl}$.

The essential difference between these two ways of treating ties is in the amount they contribute to stress. In the primary approach, the ties will contribute nothing

to stress unless they are merged with preceding or succeeding blocks by virtue of the monotone regression that is performed. The secondary approach in effect uses the mean of the distances (d_{ij}) for the tied blocks, and thus they remain tied whether or not the tied blocks are merged with others during the monotone regression. The primary approach appears to be the most widespread in the major nonmetric MDS algorithms, and a recent study has concluded that the primary approach is to be preferred [33, p. 136].

4. Goodness-of-Fit. The problem of determining the "best fitting" configuration of derived distances is also important in all the various approaches to MDS, and the method of resolving it varies from algorithm to algorithm. For example, Kruskal [28] uses the normalized residual variance from the monotone regression to define a measure (called STRESS) which forms the basis for determining the direction and magnitude of distance movements at each iteration, and which when minimized, gives an estimate of goodness-of-fit of the final configuration. Other goodness-of-fit measures include "squariance," "coefficient of alienation," and "WORK" [61, p. 9; 32, p. 489; 38, p. 185].

The index-of-fit STRESS is used both in the Kruskal MDSCAL series of nonmetric MDS algorithms and the TORSCA routines. Basically it is the normalized residual variance derived from a monotone regression of distance and dissimilarity. The monotone regression in this case involves moving only the horizontal (distance) measures in order to compute a line of best fit. This requires first matching the dissimilarity and distance measures, then checking each distance measure to see if it is greater (in a monotonically increasing relation) or smaller (in a monotonically decreasing relation) than the preceding distance. Monotonic transformations of the distance values (called disparities in the worked example in Section II) are then made to satisfy the inequality conditions in the input matrix. If a set of distance (d_{ij}) values can be generated such that, when they are arranged in vector form, they occupy the same vector position as their corresponding dissimilarity (δ_{ij}) measures, then "perfect match" is said to occur and stress is zero.

The transformed distances ($\hat{d}_{ij}$) may be regarded as the set of numbers that depart to the minimum degree from the corresponding set of computed distances, (d_{ij}), while ensuring that their rank order is the same as that of the original dissimilarities. Thus if we were to plot a curve to the scatter of ($\hat{d}_{ij}$,δ_{ij}) points, the curve should move vertically or to the right, never to the left. Since we wish to work only with the rank order of the δ_{ij}, stress focusses on the difference between the computed distances and the transformed distances, ($d_{ij} - \hat{d}_{ij}$). This ensures that any monotone distortion of the dissimilarity axis will not affect the goodness-of-fit of the configuration. Thus, if an original point is located at (d_{ij}, δ_{ij}), the point corresponding to the transformed distance has coordinates of ($\hat{d}_{ij}$,δ_{ij}). "Fitting the curve," in this case means no more than fitting M $\hat{d}_{ij}$ values (M = n(n-1)/2). Note that the $\hat{d}_{ij}$ are not distances from any configuration, but are only a monotone sequence of numbers chosen as "nearly equal" to the original d_{ij} as possible and having the interesting property that their rank order corresponds to that of the δ_{ij}.

Having determined a set of numbers monotonically related to the set of d_{ij}, define raw stress as follows:

$$\text{Raw Stress} = S^* = \sum_{i<j} (d_{ij} - \hat{d}_{ij})^2 . \quad (3)$$

Except for normalization this becomes the goodness-of-fit measure. Raw stress, while being invariant under the rigid motions of configuration (rotation, translation, and reflection), is not invariant under uniform stretching and shrinking of the configuration. To compensate for this, Kruskal originally divided raw stress by a scaling factor (T*) which was simply the sum of squared distances. Taking the square root of the final index led to STRESS being defined as:

$$S_1 = \left[\sum_{i<j} (d_{ij} - \hat{d}_{ij})^2 \; / \; \sum_{i<j} d_{ij}^2 \right]^{1/2} . \quad (4)$$

However, in later works, Kruskal decided that rather than standardize raw stress simply in relation to the sum of squared distances, it should be standardized in terms of squared deviations about a mean distance. The alternate stress value proposed, therefore, was:

$$S_2 = \left[\sum_{i<j} (d_{ij} - \hat{d}_{ij})^2 \; / \; \sum_{i<j} (d_{ij} - \bar{d}_{ij})^2 \right]^{1/2} \quad (5)$$

where $\bar{d}_{ij}$ is the mean of the distance scores.

Kruskal suggests the following verbal evaluation of goodness-of-fit:

TABLE 1. STRESS EVALUATIONS

Stress (S_2)	(S_1)	Goodness-of-Fit
40%	20%	Poor
20%	10%	Fair
10%	5%	Good
5%	2½%	Excellent
0%	0%	Perfect

Source: Kruskal [28, 29].

5. Initial configurations. Since our worked example follows the Kruskal-Shepard-Torgerson (K-S-T) mode rather than the Guttman-Lingoes smallest space analysis mode, discussion of initial configurations will be limited to the options available for the K-S-T routines.

Initial configurations may be generated entirely without bias (i.e., randomly) or may be deliberately structured in an attempt to hasten convergence of data and generated configuration. Two major alternatives are available in the Kruskal series of programs:

i) the input of a configuration "of one's choosing" (which may be arbitrarily selected);

ii) generation of a pseudo-random initial configuration.

The Young-Torgerson routine, on the other hand, uses a semi-metric method of defining an initial configuration. In this method the original dissimilarities data are transformed to scalar products (U_{ij}) using,

$$U_{ij} = \left[\left(\sum_{k}^{p} S_{ik}^2 + \sum_{k}^{p} S_{kj}^2\right) / 2p\right] - \left[\left(\sum_{k}^{p}\sum_{h}^{p} S_{kh}^2\right) / 2p^2\right] - S_{ij}^2/2 \quad (6)$$

where S_{ij} is the original similarity of points i and j, and p is the number of points [33, p. 130]. Then, assuming the "true" dimensionality of the points equals r, the *r* largest eigenroots of the Matrix U are extracted. Each vector is then multiplied by the square root of its eigenroot, and a new set of distances is calculated according to the general Minkowski formula:

$$d_{ij} = \left[\sum_{a}^{r} \left(\left| V_{ia} - V_{ja} \right|\right)^m\right]^{\frac{1}{m}} \quad (7)$$

where V_{ia} is the i^{th} scaled entry in eigenvector (a), and *m* is the Minkowski metric number selected for the problem.

Thus the original data are converted to scalar products, these scalar products are factor analyzed such that the data are produced using only the first r-dimensions, and then, finally, a monotone transformation of distances is found which best fits the original idea. Here "best" is determined by an index-of-fit. In fact, the best-fit problem is viewed as a regression problem, with a monotone regression of distances and disparities, where "distances" (d_{ij}) are the measurements produced from a given configuration by applying the Minkowski distance formula, and "disparities" are the monotonically transformed distances (referred to in our worked example by the term $\hat{d}_{ij}$).

The disparities produced by the above method then become the basis for a second factor analysis and the whole process is repeated. Thus, the Young-Torgerson routine uses a "semi-metric" algorithm to define an initial configuration prior to its manipulation by the non-metric algorithm. The routine is called semi-metric because it performs the metric operations of multiplication and addition on the original (dissimilarities) data during the first factor analysis, but uses a non-metric monotone transformation (the disparities) for each successive factor analysis.

6. Dimensionality. The final determination of the number of coordinates recovered for the data rests ultimately with the experimenter. To assist in finding an appro-

priately dimensioned solution, the following alternatives have been suggested [28].

a) Undertake the analysis in several dimensions and plot the relation between stress and dimensionality. Generally some noticeable "elbow" will occur in the curve and this should indicate the appropriate number of dimensions (Figure 1).

b) If some given t-dimensional solution, for example, provides a reasonable solution (in terms of stress size and in terms of the ability to interpret coordinates), and if a (t+1)-dimensional solution produces no major improvement in interpretability, then the t-dimensional solution should be used.

c) If there is an independent estimate of the statistical error of the data, then this gives one an idea of the appropriate number of dimensions to extract – in the sense that the more reliable the data, the more dimensions one can safely extract.

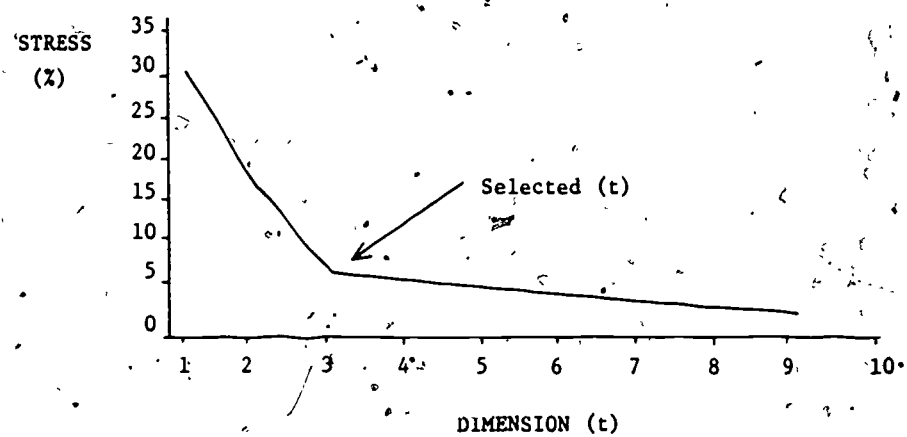


Figure 1. Shepard Diagram

7. A metric example: recovering point configurations from interpoint distances. In an unpublished paper, W. R. Tobler has developed a worked example for the metric scaling problem of finding the point configuration that most accurately reproduces a given set of interpoint distances. He has generously offered to include it here (Figure 2). Whenever the researcher is confident of the metric properties of his similarities data, the procedure described below is a far simpler and more accurate one for recovering the scale. However, when one is confident only of the ordinal relationships in the original data, non-metric scaling is more appropriate. In both cases, however, the procedure for finding point locations from distances is identical – hence the importance of this example. In non-metric scaling, the distances "fitted" are the monotonically transformed distances rather than the original distances. In metric scaling, the original distances remain at all times the target distances. Convergence is thus assured. In the case of non-metric scaling, the original

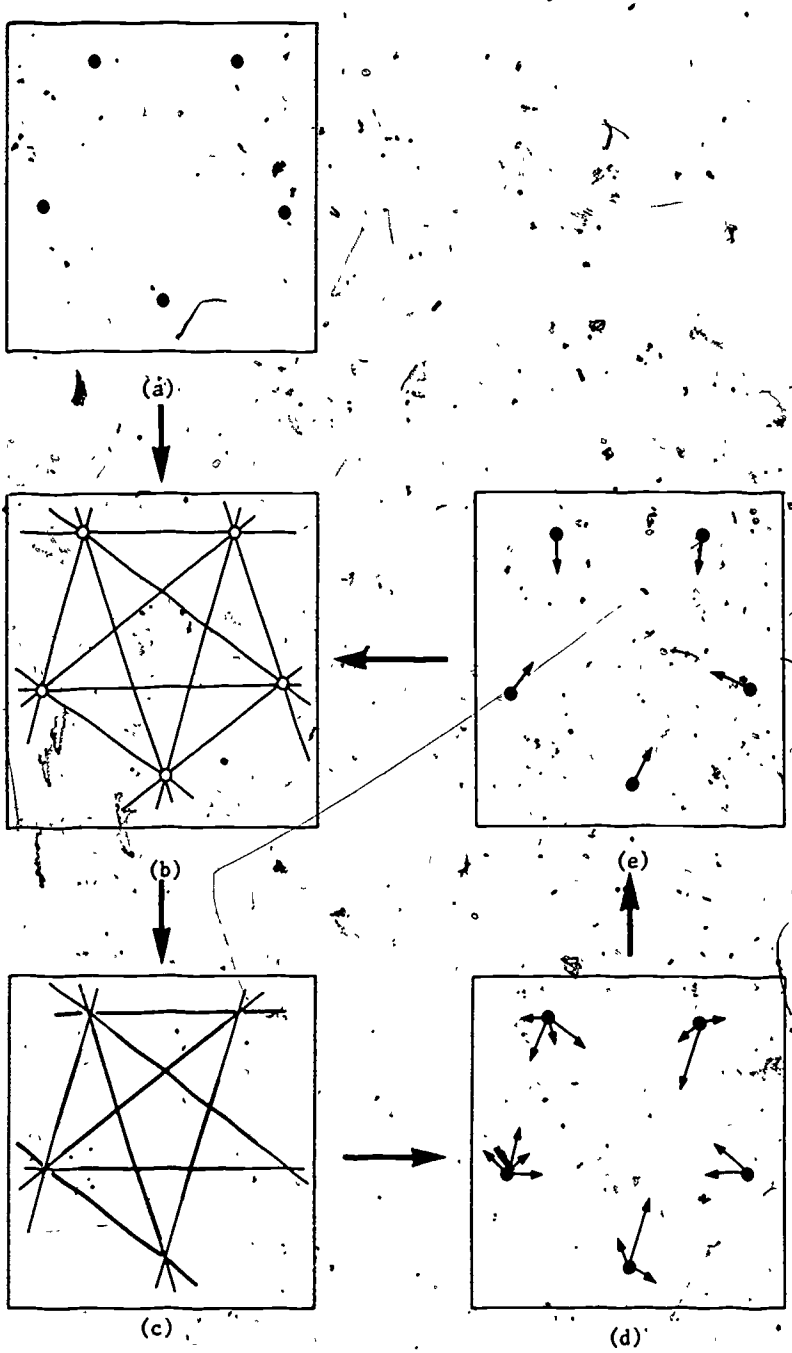


Figure 2. Trilateration Sequence (Provided by Professor W. R. Tobler, Department of Geography, University of Michigan)

distances constrain at all times the target distances via the monotonic transformation (either rank-images or monotone regression values, see below), but the actual values of the target distances change as the iterations proceed.

Tobler's triangulation problem may be stated as follows:

i) given up to n(n-1)/2 empirical distances D_{ij} between n points;

ii) find the x,y coordinates of these points in such a manner that the distances d^o_{ij} calculated from these coordinates agree as nearly as possible with the given distances D_{ij}. This means that $\Sigma(D_{ij} - d^o_{ij})^2$ is to be minimized.

An iterative graphical approximation involves the following:

Step I: Locate the *n* points arbitrarily.

Step II: Draw straight lines through each pair of points.

Step III: On each line, center a segment of the desired length. Omit this step if an observation is missing.

Step IV: Draw vectors from each point to the ends of the segments representing the desired distances.

Step V: Move each point to the new position defined by the average of the local vectors.

Step VI: If no points have moved in Step V, stop; otherwise, use the new positions to begin again at Step II.

A computational algorithm can be devised by examining the relations in the vicinity of one point in more detail (Figure 3). For each non-zero desired distance:

(1) $$\text{compute } ds_{ij} = \frac{(D_{ij} - d^o_{ij})}{2};$$

(2) compute the direction cosine of the line

$$\cos\theta_{ij} = \frac{x_i - x_j}{d^o_{ij}};$$

(3) compute the change in the x direction at the i^{th} point with respect to the j^{th} point from elementary trigonometry as:

$$dx_{ij} = \cos\theta_{ij}\, ds_{ij} = 1/2 \left(\frac{x_i - x_j}{d^o_{ij}}\right) (D_{ij} - d^o_{ij}).$$

Similarly, compute

$$dy_{ij} = \sin\theta_{ij}\, ds_{ij} = 1/2 \left(\frac{y_i - y_j}{d^o_{ij}}\right) (D_{ij} - d^o_{ij}).$$

(4) The total change in the x direction is the average of all of the partial changes:

$$dx_i = \frac{1}{n-1} \sum_{j=1}^{n} dx_{ij}$$

Similarly,

$$dy_i = \frac{1}{n-1} \sum_{j=1}^{n} dy_{ij}$$

(5) Set

$$x_i' = x_i^o + dx_i$$

$$y_i' = y_i^o + dy_i$$

$$d_{ij}' = [\,(x_i' - x_j')^2 + (y_i' - y_j')^2]\ ½$$

(6) Compare

$$\frac{1}{n-1} \sum_{j=1}^{n} (D_{ij} - d_{ij}')^2$$

with the desired accuracy level and stop, or go to step (1) using the new values.

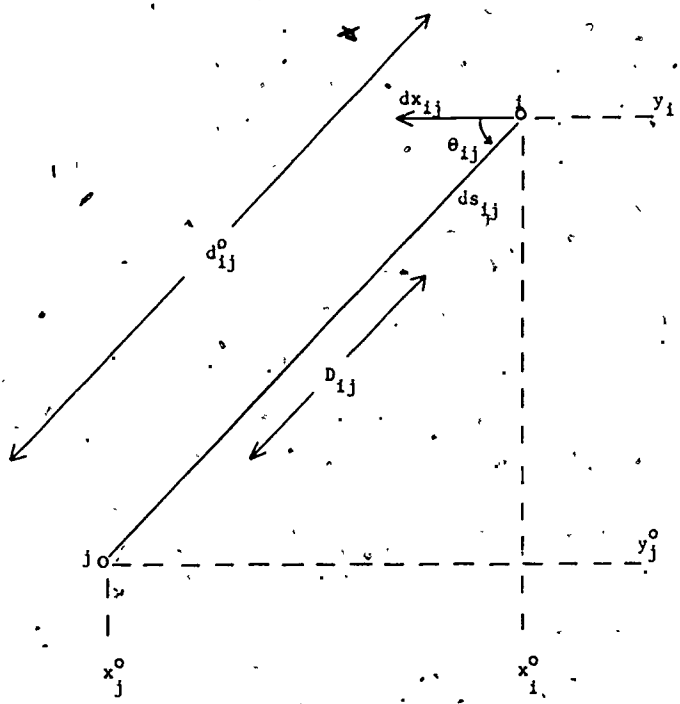


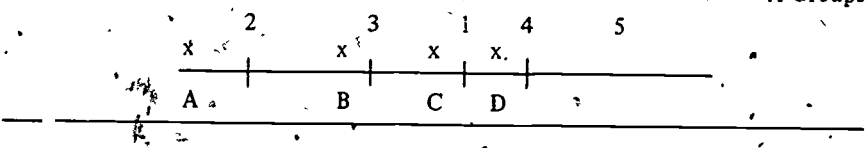
Figure 3. Trilateration Example – Explanation of Terms (Provided by Professor W R. Tobler, Department of Geography, University of Michigan)

8. **Joint-space scaling solutions.** The same principles we have already described can be applied to the problem of locating m individuals and n points in a space such that the order of the distances between the two sets of points corresponds with the original ordered data. A simple hypothetical example will illustrate the model. The rankings of five locational stimuli by four sample groups are shown in Table 2.

TABLE 2. HYPOTHETICAL RANKING OF STIMULI BY FOUR SAMPLE GROUPS

Stimuli Locational Types	Groups A	B	C	D
1	3	3	1	1
2	1	2	4	5
3	2	1	2	3
4	4	4	3	2
5	5	5	5	4

Geometrical Model:



Legend:
'1 Stimuli
X Groups

The distances from each of the groups to each of the five stimuli, when rank ordered for each group, have the same order as in Table 1. The geometrical model is useful because it summarizes the data in Table 1 ($n + m$ coordinates are sufficient to recover $n \times m$ original data values), and because it also allows generalizations to be made about similarity between groups as they order the stimuli. For example, distances between the groups in the model space may be used as a summary of similarity of point of view.

D. Selected Approaches to Non-Metric Multidimensional Scaling Analysis

There are at present a number of closely related approaches to the problem of multidimensional scaling (MDS), and some of these varied approaches are reviewed below.

1. **Coombs.** This MDS model is based on the 1954 work of Hays and is adapted from the multidimensional unfolding of preferential choice data [7, pp. 444-462]. The basic requirement for this model is that at least a partial ordering of the interpoint distances between n points can be obtained. The intent is to put a frame of reference consisting of r axes on the space in which the points are located, and to determine the rank order of the projection of all points on these lines. Every pair of

points in space defines a line and each line is a potential axis of the space. The aim of the technique is to select a minimal subset of the lines and determine the order of projections on each in such a manner as to satisfy a given partial order of interpoint distances. The criterion for selecting lines is a least squares fit to the space.

The fundamental feature of this approach is the finding of a partial ordering of interpoint distances. For any set of individuals, a complete (or simple) order of interpoint distances between objects is said to occur when exactly similar rankings of the objects are obtained. Assume there are only two individuals (X,Y), who perform a paired comparison experiment on five objects (A, B, C, D, E) and come up with the following orderings of data:

For X: EBCDA
For Y: ADCBE

These rankings indicate complete order in the data. They can be plotted in a space between individuals X and Y such that the interpoint relations for each pair of distances is satisfied. Thus, X-E-B-C-D-A-Y. Here the pair EA is defined as being most dissimilar by both X and Y and all other distance relations (such as BC, DA, etc.) can be maintained in this ordering. If however the order of data for X and Y was as follows:

For X: EBCDA
For Y: AECDB

then complete order is not obtained and only partial order is derived. Here C, and D can be located between B and A for both individuals, but their exact order and the position of E is not easily determined. For example, all we can deduce from this is that E falls between B and A for Y but not for X (Figure 4).

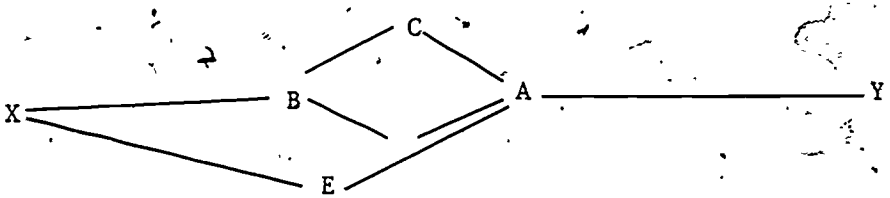


Figure 4. Partial Order (After Coombs, 1964)

The problem attacked by Coombs, then, is that of resolving partial orders into simple order by searching for the minimal number of dimensions required to locate these objects. When projected onto the selected dimensions, the distances between points located in the resulting configuration must reflect their position and relations in the orderings given by individuals.

The methodology involved is an iterative procedure based on least squares principles. The first step is to check for unidimensionality by selecting the largest interpoint distance and seeing if a simple order of objects is obtained. If not, partial orderings are examined to see how many objects can be located correctly with

respect to this dimension and incomparable relations (such as the C-D order in the example above) are resolved where possible. A second dimension is then chosen from the remaining incomparables, and another attempt made to produce a simple order of these remaining distances. Success would indicate a two-dimensional solution to the original problem. failure involves selection of a third and perhaps other dimensions using the same principles of selection. Final axes are generally "reasonably orthogonal" to each other. The end result is a configuration of points in an r-dimensional space.

2. Torgerson. Torgerson's original methodology appeared to have much in common with factor analysis and, in fact, used a centroid method of factor analysis to derive the factor scores which were used to define the final configuration of object-points. Taking as input triadic comparisons data, Torgerson converted initial similarity proportions into standard (Z) scores. These scores were then transformed into a single matrix of "comparative distances" between objects and an attempt was made to find an "additive constant" which translated the comparative distances to absolute distances. In other words, the absolute distance between objects (d_{ij}) was regarded as being a combination of some perceived or psychological distance (h_{ij}) and the additive constant (C).

Torgerson [57] suggested that both the formal mathematical method of Messick and Abelson and his own shortcut method could be used to define the additive constant in any given experimental situation. Once a matrix of absolute distances had been compiled, it was transformed again to a scalar product matrix (B*) and referred to an origin at the centroid of all the stimuli. The B* matrix was then factored (using centroid procedures), the relevant factors examined, and factor scores used as coordinates on each dimension in order to specify the nature of the final configuration of points. The essence of this approach, therefore, was to find the kind of distance function that was necessary to convert (psychological) measures of similarity into real-number measures.

The differences between this early approach and the one examined in detail in this paper result from the adoption of a weaker assumption (the monotonicity assumption) in transforming psychological to real distance measures. While some basic similarities are retained, the emphasis in the TORSCA program used here is not one of finding an additive constant but rather one of developing a set of monotonically related measures which correspond to the original dissimilarity measures. Since our detailed example breaks down the TORSCA algorithm into each of its component parts, no further expansion is necessary here.

3. Shepard. Unlike much of the earlier work in MDS, Shepard [50] did not try to deal explicitly with a definition of the distance function required to translate "psychological" into "real" distances. His aim was to find a configuration of points in a minimum number of dimensions such that a plot of psychological distances against real distance would reveal the specific function needed for the translation. His technique was to allocate a set of (n-1) vectors to each of n-stimulus points

directed to each of the other (n-1) stimuli, and to indicate at each iteration how the stimulus point should be moved in order to improve the monotonicity requirement. At each iteration there was simultaneous displacement of all these points.

In order to obtain a minimum dimensional configuration, however, another set of (n-1) vectors was defined for every (n) point which aimed at increasing the variance of distances at each iteration. By aiming at an increase in the variance of distance it was possible to collapse the configuration into smaller spaces. To effect this increase in variance, large distances were made larger, and small distances made smaller. The iterative procedure was halted when an index (S) defined as:

$$S = \left[\sum_{i,j} (S_{ij} - S(d_{ij}))^2 \; / \; n(n-1) \right]^{1/2} \qquad (8)$$

where S_{ij} = proximity measure between S_i and S_j
$S(d_{ij})$ = proximity measure corresponding to the computed distance (d_{ij}),
and n = number of stimuli,

became "small enough". Since the stimuli would be still defined in terms of coordinates in an (n-1) space, in order to expose the minimum dimensional configuration he had (like Torgerson) to determine a scalar product matrix, obtain a roots and vector solution to the matrix, and eliminate "unimportant" axes.

4. Kruskal. Shepard's focus on monotone relationships rather than a specific mathematical function designed to translate psychological into real distances provided the impetus for Kruskal's approach to multidimensional scaling [28, 29, 30]. His basic problem was to find a set of distance measures which could be related to the dissimilarity measures such that monotonicity is not violated, and such that a monotone regression of dissimilarity and distance yields a minimal value for STRESS.

After generating an arbitrary initial configuration which is used to calculate distances and to generate disparities, Kruskal produces convergence of the configuration and original data by a method known as the method of steepest descent (or method of gradients). First the gradient of raw stress is determined for each point in the configuration in the space of specified dimensionality:

$$g_{ia}(t) = \sum_{j=1}^{n} \left[\left(1 - \frac{\hat{d}_{ij}(t)}{d_{ij}(t)} \right) \cdot \left(x_{ja}(t) - x_{ia}(t) \right) \right] \qquad (9)$$

where g_{ia} is the gradient of stress for distance (i) at location (a),
$\hat{d}_{ij}(t)$ is the estimate of distance at iteration (t),
$d_{ij}(t)$ is the actual distance measure at iteration (t),
and x_{ja} & x_{ia} are orthogonal coordinates of points i and j.

The gradient provides directional information for moving the points in the configuration along the path of steepest descent such that $S^{(t+1)} < S^{(t)}$ (S = Stress). In essence, the gradient is determined by taking the first partial derivative of stress with respect to each point (assuming the partials are negative) [29, p. 118],

$$\frac{-\delta S}{\delta x_{1t}}, \ldots\ldots\ldots, \frac{-\delta S}{\delta x_{it}}, \ldots\ldots\ldots, \frac{-\delta S}{\delta x_{nt}} \qquad (10)$$

Once the directions of movement for points in the configuration have been determined, the distance that each is moved (step size) is determined, using:

i) an estimate of the previous step size (α) – for the first movement α can be arbitrarily set (e.g., $\alpha = 0.2$) [21, p. 121];

ii) an angle factor (f_1) which represents the cosine of the angle between successive gradients in successive dimensions (defined as $4.0\ (\text{Cos}\theta)^3$ where θ is the angle between successive gradients);

iii) a relaxation factor (f_2) which is defined as:

$$f_2 = \frac{1.3}{1 + (\text{5-step-ratio})^{5.0}}$$

where the five-step-ratio $= \min\left[1, \left(\frac{\text{present stress}}{\text{stress 5 iterations ago}}\right)\right]$;

iv) a "good luck" factor (f_3) defined as:

$$f_3 = \min\left[1, \left(\frac{\text{present stress}}{\text{previous stress}}\right)\right].$$

Note in ii) aboye, if g represents the present gradient and g'' the previous one, then [29, p. 122]:

$$\cos\theta = \frac{\sum_{i,s} g_{is}\, g''_{is}}{\left(\sum_{i,s} g_{is}^2\right)^{1/2} \left(\sum_{i,s} g''^{2}_{is}\right)^{1/2}} \qquad (11)$$

Thus step size at iteration (t) is defined as:

$$\alpha(t) = \alpha(t\text{-}1), f_1, f_2, f_3 .$$

This provides for relatively large steps at the beginning of the iterative process and fairly small steps towards the end. Although the length of gradient is partly used as a source of information for termination of the iterative process, Kruskal relies primarily on the angle between successive gradients to guide the iterative process. The calculation of step size in terms of STRESS helps to partly overcome the

problem of local minima, for if STRESS is large the iterative procedure will continue, while for low STRESS the local minima may prove a "satisfactory" configuration.

The outcome of manipulating distance estimates is to obtain a configuration of points in a space of specified dimensionality in which the distance relations conform to the original dissimilarity measures. Like factor analysis, the problem may then arise as to how to interpret the dimensions of the configuration, but at least one is aware of the number of attributes being used by a population in order to make comparative judgments concerning the original stimuli.

5. Guttman-Lingoes. These multidimensional scaling programs known as the Smallest Space Analysis (SSA) series are developed from the most comprehensive algebraic treatment of multidimensional scaling to date [18; 33].

The starting configuration is similar to that in TORSCA described above except that it is the rank order of the input similarity measures, rather than the s_{ij} values themselves, that are operated on and the initial factoring of the rank order matrix is not an iterative procedure as in the TORSCA algorithm. The loss function (ϕ) (cf. Kruskal's stress), is defined as:

$$\phi^* = \sum_{i=1}^{n} (d_i - d_i^*)^2 \quad : \quad (n=1/2n(n-1))$$

and the normalized phi:

$$\phi = \sum_{i=1}^{n} (d_i - d_i^*)^2 \; / \; \sum_{i=1}^{n} d_i^2 \qquad (12)$$

where d_i^* are the monotonically transformed values of d_i know[illegible] the "rank images" of the d_i [18, p. 479; 33, p. 9]. These rank images are obta[illegible]ing the d_i and by placing them in the cells corresponding to the cells ra[illegible]ered on the input similarity values (s_{ij}). Thus, the smallest interpoint dista[illegible] computed from a given configuration becomes the rank-image of the interp[illegible]t distances corresponding to the smallest s_{ij} value . . . and so on.

The d_i^* serve two purposes: first, they provide target estimates t[illegible] since their rank-order corresponds to that of the input data, may be used to gu[illegible] subsequent iterations toward a solution configuration; and second, they provide [illegible]lues that can be used to measure progress in reaching a final solution. Thus, whe[illegible] $\sum_{i=1} (d_i - d_i^*)^2$ = 0, perfect fit obtains. The measure of unexplained variance (the "coefficient of alienation") is defined:

$$K = [1 - (1 - \phi)^2]^{1/2}. \qquad (13)$$

The method of minimizing K is known as the "Correction Matrix" meth[illegible] is similar to that used in the non-metric algorithm in the TORSCA routine (descr[illegible]

in detail in the worked example in Section II), with modifications in the step-size adjustment parameter (α). In the SSA series, alpha varies not only from one iteration to another (as in both the MDSCAL and TORSCA algorithms) but also from one point to another. A further refinement in the SSA series is that iterations are allowed to proceed for a given set of d_i^* before the re-definition of a new set of d_i^* on the basis of new computed values of d_i, (known as the two-phase process). Some advantages are derived by formulating the algorithm in these terms [18, pp. 485-486].

E. Selected Problems in Non-Metric Scaling

We discuss below two major problem areas which must be confronted in using multidimensional scaling. One is an interpretive problem—that of finding substantive meaning in the dimensions of the recovered configurations. The other is a problem in algorithm construction and is concerned with ensuring that whatever goodness-of-fit function the algorithm has attempted to minimize has in fact been minimized. Some of the earlier scaling programs were particularly prone, under certain conditions, to lead to "solutions" that were far from the optimal possible. This problem of local minima and convergence is discussed first.

1. Local minima and the convergence process. A goal of all of the scaling algorithms discussed was to avoid situations in which points would become located such that small movements, however computed, would always lead to higher stress even though some other configuration of the points might exist for which stress would be appreciably lower. A number of strategies have been suggested to ensure that the researcher will not present such interim solutions as final ones.

a. Multiple solutions. In this strategy [25, p. 158; 33, p. 117] the researcher computes a number of solutions from different (usually random) starting configurations. Thus the hope is that local entrapment of points will be avoided in at least some of the "runs." With this strategy, the best-fitting solution configuration is accepted [20].

b. Selection of initial configuration. Evidence exists [33, pp. 110-113, 126-132], [53] that starting configurations influence the likelihood of encountering local minima solutions. Random or arbitrary configurations in particular are prone to this problem [33, p. 126]. Sometimes, the researcher may have knowledge about some of the properties of the final configuration and in such cases he may wish to use a configuration containing these properties as the starting configuration. Young goes and Roskam [33, p. 127] recommend analysis of the input similarities to compute a starting configuration for the non-metric scaling algorithm.

c. Backup procedures to confirm convergence. In later versions of Kruskal's MDSCAL algorithm [28], the technique employed to mitigate the occurrence of

local minima problems is to retain after any iteration the capability of returning to the configuration prior to that iteration if either stress has increased by 20% or more, or if the angle of the gradient after the iteration (computed to determine the direction of movement for any point) is close to being directly opposite (180°) to its value prior to the iteration.

d. Varying the target distances. The distances that the configuration points are attempting to match are of crucial significance to the local minima problem. While the d_{ij} of Kruskal have the advantages of speedily converging (when used with a sensitive step-size adjustment procedure), they have the disadvantage of changing in value between iterations. With both configuration distances and target distances changing, smooth convergence is not guaranteed. Indeed, erratic behavior sometimes occurs. Lingoes and Roskam [33, pp. 133-136] have experimented with an algorithm in which the target distances are, alternatively, the d* (rank-image distances used to minimize the Guttman-Lingoes coefficient of alienation) and the $\hat{d}$ (monotone-regression distances or disparities) used to minimize stress. This sequence of targets applied to experimental data in a succession of analyses showed very low incidences of local minima. They conclude that, although this strategy involves more iterations than one which stayed with the one set of target distances, it is sufficiently impressive in avoiding local minima problems that it is incorporated in their revised algorithm MINISSA.

2. **Interpretations of scale dimensions.** One of the fundamental problems in the use of MDS is that of interpreting the dimensions in which configurations are mapped. In some cases this particular problem does not arise because the researcher is interested only in the position or relation of the configuration points with respect to each other (for example, in determining clusters of stimulus points). However, in other cases the scale value derived from projecting any given point onto an appropriate dimension of the space is sought after. Under these circumstances the interpretability problem arises.

The problem of identifying the dimensions of any configuration has been resolved in a number of different ways. For many of the geographical studies mentioned in this paper, some information is known about the location of stimulus points in an objective space prior to the building of a configuration. When this type of information is known, constraints can be placed on the number of dimensions in which the output configuration is produced such that one attempts to replicate the spatial structure of the original objective configuration. Thus in Tobler's use of MDS algorithms as map transformations, he is able to orient configurations in the same way as preselected map projections are oriented, and to interpret distances according to distances measured on these projections. Similarly, in the urban distance perception studies discussed later, the locations of the phenomena being investigated were known prior to the beginning of the study. Output configurations could therefore be rotated, reflected, and translated until the positional relations in the output configuration have the same directional components as in the original

data. In both of these cases, the problem of identifying the dimensions themselves are trivial ones.

In cases where the configuration of stimulus points is not known, considerable ingenuity has been used in order to interpret dimensions. For example, in one study (III-A-2) a large number of personal, social, economic, and attitudinal characteristics were collected for each sample respondent, and various types of characteristics were collected for the stimulus points. By investigating the scores of stimulus points and individuals on unidimensional scales of each attribute, the author was able to interpret his dimensions by choosing those attributes which appeared to be most highly correlated with the derived scale values. In another study (III-A-3), dimensions of the configuration were interpreted in terms of qualities of the stimulus objects that had been derived from independent scaling analysis. Thus, when the configuration of stimulus points is unknown, it appears that the most appropriate method for identifying dimensions is to compare the scores of each stimulus point on each dimension with some prior selection of attributes of the stimuli. Those attributes having the highest correlation with scale values then lend themselves to use in interpreting dimensions of the configuration.

Another point which is seldom discussed in the literature is that of the positioning of the axis of the configuration space. It appears that, for the most part, scale values are referred to the centroid of various configurations. This means in effect that an arbitrary zero point is established and the data can be interpreted at no higher level than an interval scale. In terms of the problems associated with interpreting configurations, therefore, there are certain similarities with other multivariate techniques such as principal components and factor analysis. We may conclude that there are corresponding difficulties in the interpretation of dimensions just as there are in these other metric methods of multivariate analysis.

II. THE NON-METRIC ALGORITHM: A WORKED EXAMPLE

Our discussion to this point has centered on the objectives behind multidimensional scaling approaches and on the question of the appropriate criteria for evaluating goodness-of-fit. Several algorithms have been developed to meet the stated objectives, and only recently have any systematic attempts been made to compare the relative performances of the different algorithms [33, 43, 53, 62]. It is our purpose here to explain multidimensional scaling through the mechanics of a hand-worked example, and we choose in the example below to solve a problem using one of these several scaling algorithms.

The particular one chosen was based on the availability, at the time of writing, of a clear description of the algorithm. The one used is that developed by Young and Torgerson [61, 63], and it is applied below to a very simple, hypothetical problem. In the following section of this paper we will discuss more realistic examples. However, all such applications of multidimensional scaling techniques to date have been with data containing too many points for solving in a hand-worked example in a reasonable length of time. Even the simple hypothetical problems described below would take a day or so to solve by hand-calculator methods.

A. The First Example

The problem we outline is one in which the solution is known at the outset, but the worked example below is such that the information derived as a solution is not explicitly contained in any of the initial data from which it proceeds. The multidimensional scaling problem recall is essentially one of finding the locations of points in a space of any given number of dimensions such that an ordering of the distances between points in this space best corresponds to an ordering of the points in the original input data. Thus, it is implied that the researcher possesses knowledge on the order relations between a set of points (usually from experimental data) and that he hypothesizes these order relations are derived from a mental configuration of the points (unknown, of course, to the researcher). The purpose of the scaling, then, is to construct a configuration of points from which measurements can be made between points on which the order relation corresponds to the order relation of the interpoint distances in the experimental data.

In Figure 5 we portray an arbitrary configuration of four points in two dimensions, in Table 4 the distances between the points are shown; and in Table 5 the order relation of these distances is shown. In the worked example, the information of Table 5 is the only part of this initial data that is used. However, since it is derived from Table 3 which in turn is derived from Figure 5, it is possible to check the accuracy of the solution developed below by comparing it with the arbitrary configuration in Figure 5. In brief, therefore, the problem can be described as that of deriving the essential properties of this figure solely from the information contained in Table 5.

TABLE 3. COORDINATES FOR THE ARBITRARY CONFIGURATION IN FIGURE 5

	Axis a	Axis b
Point 1	1.0	1.7
Point 2	1.1	1.2
Point 3	2.05	2.0
Point 4	2.15	0.4

TABLE 4. INTERPOINT DISTANCES IN THE ARBITRARY CONFIGURATION

	1	2	3	4
1	0.0	0.5	1.09	1.75
2	0.5	0.0	1.25	1.30
3	1.09	1.25	0.0	1.60
4	1.75	1.30	1.60	0.0

TABLE 5. ORDER-RELATION OF DISTANCES IN TABLE 4

	1	2	3	4
1	0.0	1.00	2.00	6.00
2	1.00	0.0	3.00	4.00
3	2.00	3.00	0.0	5.00
4	6.00	4.00	5.00	0.0

1. Overall strategy. The problem is solved in a series of interations, each one of which comprises four stages. We begin with a random configuration of the four points in two dimensions[1] (the initial configuration).

The strategy for solution is to move successively closer to the solution on each iteration, stopping when the index of fit shows that the previous iteration has resulted in a new configuration that is not superior to that of the previous one.

2. The four stages. Step 1. *Computation of distances:* Distances between the points of any configuration are computed from the formula:

$$d_{ij} = \left[\sum_{a}^{t} \left(\left| x_{ia} - x_{ja} \right| \right)^{m} \right]^{1/m} \qquad (14)$$

[1]The worked example does not exactly parallel the method suggested by Young and Torgerson [63] since, in the interest of reducing the total number of iterations required to reach a solution, they suggest a strategy of first preparing an initial configuration by assuming metric properties in the input data and then optimizing the configuration by the method outlined below. In the first section of this example we confine ourselves to a discussion of the nonmetric algorithm since this is the essential step common to all multidimensional scaling techniques. The computation of initial configurations with properties approaching final configurations is discussed later in this section.

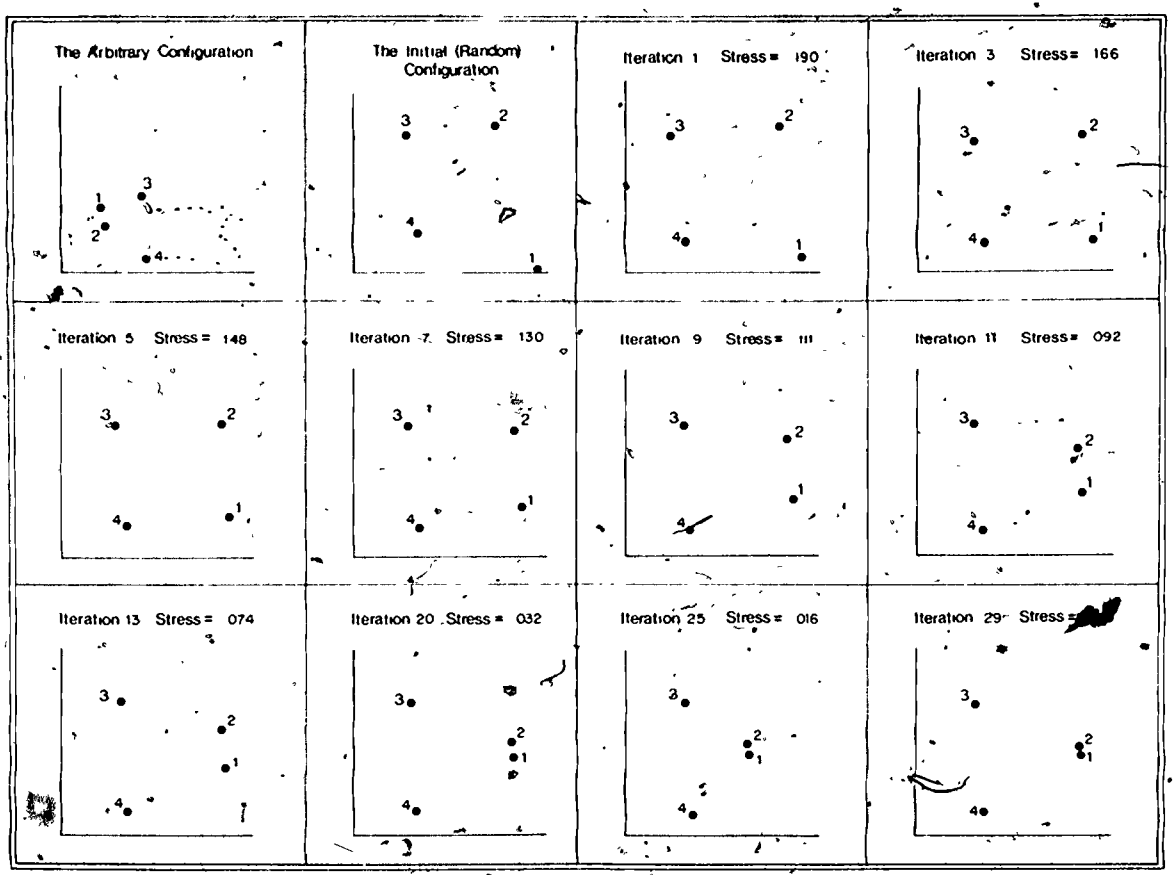


Figure 5. Progress of the Iterations for the First Example

where i and j represent any two points in the configuration, x_{ia} and x_{ja} are the coordinates of the points on axis a, r is the number of axes (dimensions) for which the solution is being computed, and m is the Minkowski constant [52] determining the type of distance metric for which the solution is being determined, e.g., Euclidean distance has the Minkowski constant of 2.

Step II. *Computation of monotonically transformed distances (disparities):* The purpose of this step is to constrain the distances from the given configuration so that they do not violate the order relation of the original interpoint distances (in this case, Table 5). This is achieved by ensuring that if one were to plot the values of this table (y axis) against the monotonically transformed values of Table 7 (x axis), the "curve" joining these points would never move to the left but only vertically or to the right. These monotonically transformed distances are known as "disparities"; they are not distances from any known configuration but rather are a monotone sequence of numbers as "nearly equal" to the distances in the given configuration as is possible without violating the original order relation of interpoint distances. It is reiterated that the purpose of non-metric multidimensional scaling is to construct a configuration of points in any given number of dimensions such that the interpoint distances and the monotonically constrained distances (disparities) are as similar as is possible [28, 29, 61].

TABLE 6. COORDINATES FOR THE INITIAL (RANDOM) CONFIGURATION IN FIGURE 5

	Axis a	Axis b
Point 1	4.70	0.10
Point 2	3.60	3.90
Point 3	1.30	3.60
Point 4	1.70	1.00

TABLE 7. INTERPOINT DISTANCES IN THE INITIAL CONFIGURATION

	1	2	3	4
1	0.0	3.95	4.88	3.14
2	3.95	0.0	2.31	3.47
3	4.88	2.31	0.0	2.63
4	3.14	3.47	2.63	0.0

The disparities are computed by taking the interpoint distance from the configuration (Table 7) corresponding to the two most similar points (smallest distance) in the original data (Table 5) and comparing it with the distance that corresponds with the next two most similar points. If the first distance is smaller than the second, then the order of the distances corresponds with the order of the original similarities and no transformation is necessary. However, if the first distance is larger than the second, then the original order relation of interpoint distances is violated; since the disparities must not decrease in value when the similarities increase, the arithmetic mean of the two distances is substituted for the

distances and this mean becomes the first two disparity values. The interpoint distance from the configuration corresponding to the third smallest original similarity (Table 5) is then compared with the second disparity (which might be the second distance or the mean of the first two distances—as discussed above). If the third distance is larger than the second disparity, it becomes the third disparity; otherwise, the mean of it and the previous disparity (weighted mean if the previous disparity was composed of more than one distance) is computed, and this mean becomes the disparity value for the third distance and second distance (also for the first distance if the second disparity was itself a mean value). In the first iteration of the problem described below, the disparities will be computed step by step following the procedure outlined above.

Step III. *Computation of goodness-of-fit (stress):* The measure of goodness-of-fit is a measure of how far the disparities ($\hat{d}_{ij}$) depart from the distances measured from the derived configuration (d_{ij}). The larger these departures are, as compared with the distances themselves, the greater the error in reproducing the order relation of the original similarities from the derived configuration and therefore the poorer the fit. Kruskal's "stress" values as defined in equations (4) and (5) are both computed in the example outlined here. The smaller the stress, the better the fit.

Step IV. *Computation of a new (improved) configuration:* As noted, the greater the discrepancy between the distances and disparities from any configuration, the poorer the configuration. Therefore, to improve any given configuration each point should be moved so as to reduce the average discrepancy between the distances and the disparities with respect to the other points [29, pp. 117–123; 61, pp. 6–7]. If $d_{ij} > \hat{d}_{ij}$ then point i should be moved closer to point j by an amount proportional to the size of the discrepancy. Thus, *ceteris paribus*, after this adjustment the discrepancy $d_{ij} - \hat{d}_{ij}$ should be smaller on the new configuration than on the previous one. However, since for each of n points there are $n - 1$ distances to the other points, there will be $n - 1$ possible adjustments for each point. The mean of these possible adjustments is the actual adjustment that is made. In the worked example that follows, the formulae used are from Young [61, p. 6]. The displacement of point j with respect to point j is given by:

$$c_{ija} = \alpha (d_{ij} - \hat{d}_{ij}) \cdot (x_{ja} - x_{ia}) / d_{ij} \qquad (15)$$

where d_{ij} are distances computed from the previous configuration,

$\hat{d}_{ij}$ are the disparities (monotonically transformed distances) from the previous configuration,

x_{ja} and x_{ia} are the coordinates of points i and j respectively on axis a,

and α is a constant of proportionality.

The new position of point i on axis a is the coordinate on the previous configuration plus the means of the correction vectors defined in (15) above:

$$c_{ia} = \frac{\alpha}{n} \sum_{j=1}^{n} (d_{ij} - \hat{d}_{ij}) \cdot (x_{ja} - x_{ia}) / d_{ij} \qquad (16)$$

$$x_{ia}^{1} = x_{ia} + c_{ia} \qquad (17)$$

Alternative formulae for reaching the same goal are discussed elsewhere, [29, pp. 117–126; 18, pp. 484–500; 43, pp. 23–29].

In the literature of multidimensional scaling, alpha is commonly referred to as the "step-size parameter." Much discussion has centered on the question of modifying the value of alpha as the iterations converge on a solution [18, pp. 491–492, 29, pp. 121–123; 43, pp. 23–29]. In the method of computing revised configurations used in this example, larger values of alpha would speed up convergence (that is, lead to a given solution in a small number of iterations) but subject to the constraint that if alpha becomes too large, there is the danger of moving the points in any revised configuration too far. In this example, for the sake of simplicity, alpha is given the value of 1.0.

3. **Application of the four stages to the sample problem.** Step I. *Compute the distances:* Applying formula (14) to the coordinates of the initial configuration (Table 6) yields the information in Table 7:

$$\text{e.g., } d_{23} = [(x_{2a} - x_{3a})^2 + (x_{2b} - x_{3b})^2]^{1/2}$$

$$= [(3.60 - 1.30)^2 + (3.90 - 3.60)^2]^{1/2}$$

$$= 2.31$$

Step II. *Compute the disparities:* Following the order relations of the distances in the original arbitrary configuration (Table 5), the distances in Table 7 must be transformed so that their order does not violate the order of Table 5. This is achieved in a series of trials as set out in Table 8 below. In Trial 1 the distances in the initial configuration corresponding to the first two distances in the original data are seen to satisfy the constraint of ascending order. However, the third distance (2.31 in Table 7) violates this order, and therefore in Trial 2 it is combined with the previous distance (4.88) to produce a mean of 3.60. Since this new number is smaller than 3.95 (the first distance) it must be combined to produce a new disparity of 3.71 (Trial 3). In such a manner, the final set of disparities are computed in Trial 6 where all have the value of 3.39.

Step III. *Compute stress:* Applying formula (4) to the distance-values in Table 7 and the disparity values on the final trial in Table 8 yields:

$$S_1 = \left[\frac{\sum_{i<j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i<j} d_{ij}^2} \right]^{1/2}$$

$$= \left(\frac{4.347}{73.571}\right)^{1/2} = \sqrt{.05908} = 0.234; \ (S_2 = 1.0).$$

TABLE 8 COMPUTATION OF DISPARITIES IN THE INITIAL CONFIGURATION

Rank	Interpoint order in original data (Table 5)	Corresponding distances in initial configuration (d_{ij})	Disparities ($\hat{d}_{ij}$) Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6
1	1, 2	3.95	3.95		3.71	3.65	3.45	3.39
2	1, 3	4.88	4.88	3.60*	3.71	3.65	3.45	3.39
3	2, 3	2.31	2.31*	3.60*	3.71	3.65	3.45	3.39
4	2, 4	3.47			3.47*	3.65	3.45	3.39
5	3, 4	2.63				2.63*	3.45	3.39
6	1, 4	3.14					3.14*	3.39

*indicates rank-order violated

Step IV. *Compute a new configuration.* Applying formulae (15) and (16) and substituting the distances from Table 7 and the disparities from Table 8, the adjustments to the coordinates on both axes of the initial configuration are computed below.

ITERATION 1

Point 1 on Axis 1

(3.95 − 3.39) × (3.60 − 4.70)/3.95 = −0.16 — from formula (16)
(4.88 − 3.39) × (1.30 − 4.70)/4.88 = −1.04
(3.14 − 3.39) × (1.70 − 4.70)/3.14 = 0.24

−0.24 is mean move (alpha = 1.00) — from formula (17)

Point 2 on Axis 1

(3.95 − 3.39) × (4.70 − 3.60)/3.95 = 0.16
(2.31 − 3.39) × (1.30 − 3.60)/2.31 = 1.08
(3.47 − 3.39) × (1.70 − 3.60)/3.47 = −0.04

0.30 is mean move (alpha = 1.00) — from formula (17)

Point 3 on Axis 1

(4.88 − 3.39) × (4.70 − 1.30)/4.88 = 1.04
(2.31 − 3.39) × (3.60 − 1.30)/2.31 = −1.08
(2.63 − 3.39) × (1.70 − 1.30)/2.63 = −0.12

−0.04 is mean move (alpha = 1.00)

Point 4 on Axis 1

(3.14 − 3.39) × (4.70 − 1.70)/3.14 = −0.24
(3.47 − 3.39) × (3.60 − 1.70)/3.47 = 0.04
(2.63 − 3.39) × (1.30 − 1.70)/2.63 = 0.12

−0.02 is mean move (alpha = 1.00)

Point 1 on Axis 2

(3.95 − 3.39) × (3.90 − 0.10)/3.95 = 0.54
(4.88 − 3.39) × (3.60 − 0.10)/4.88 = 1.07
(3.14 − 3.39) × (1.00 − 0.10)/3.14 = −0.07

0.38 is mean move (alpha = 1.00)

Point 2 on Axis 2

(3.95 − 3.39) × (0.10 − 3.90)/3.95 = −0.54
(2.31 − 3.39) × (3.60 − 3.90)/2.31 = 0.14
(3.47 − 3.39) × (1.00 − 3.90)/3.47 = −0.07

−0.12 is mean move (alpha = 1.00)

Point 3 on Axis 2

(4.88 − 3.39) × (0.10 − 3.60)/4.88 = −1.07
(2.31 − 3.39) × (3.90 − 3.60)/2.31 = −0.14
(2.63 − 3.39) × (1.00 − 3.60)/2.63 = 0.75

−0.11 is mean move (alpha = 1.00)

Point 4 on Axis 2

(3.14 − 3.39) × (0.10 − 1.00)/3.14 = 0.07
(3.47 − 3.39) × (3.90 − 1.00)/3.47 = 0.07
(2.63 − 3.39) × (3.60 − 1.00)/2.63 = 0.75

−0.15 is mean move (alpha = 1.00)

The new coordinates are computed by adding the mean move of any point on any axis to the prior position of that point on that axis. Thus, from Table 6 and the computations above a new set of coordinates is calculated:

Point 1 on Axis 1: 4.70 + −.24 = 4.461
Point 1 on Axis 2: 0.10 + 0.38 = 0.484
Point 2 on Axis 1: 3.60 + 0.30 = 3.897
Point 2 on Axis 2: 3.90 + −.12 = 3.784

Point 4 on Axis 2: 1.00 + −.15 = 0.847

**Missing information may be calculated from information previously given.

TABLE 9. CONFIGURATION

	1	2
1	4.461	0.484
2	3.897	3.784
3	1.262	3.486
4	1.680	0.847

Step V. *Compute the new distances (i.e., repeat Step I).* The new interpoint distances can be calculated from formula (14) and from the coordinates given above:

$$\text{e.g., } d_{12} = [(4.461-3.897)^2 + (0.484-3.784)^2]^{1/2}$$

$$= \sqrt{11.2081}$$

$$= 3.348$$

Remaining distances can be calculated from data in Table 9, above. They will have the values in Table 10, Below:

TABLE 10. DISTANCES

	1	2	3	4
1	0.0	3.348	4.387	2.805
2	3.348	0.0	2.652	3.680
3	4.387	2.652	0.0	2.672
4	2.805	3.680	2.672	0.0

Step VI. *Compute the new disparities (i.e., repeat Step II).* These new values are given in Table 11.

Order of interpoint distances in original data	Distances in the new configuration	Disparities Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
1,2	3.348	3.348	3.348	3.348	3.348	3.257
1,3	4.387	4.387	3.520	3.520	3.348	3.257
2,3	2.652	2.652*	3.520	3.520	3.348	3.257
2,4	3.680		3.680	3.176*	3.348	3.257
3,4	2.672		2.672*	3.176*	3.348	3.257
1,4	2.805				3.348	3.257
					2.805*	3.257

*order relation violated

TABLE 11. DISPARITIES

	1	2	3	4
1	0.0	3.257	3.257	3.257
2	3.257	0.0	3.257	3.257
3	3.257	3.257	0.0	3.257
4	3.257	3.257	3.257	0.0

Step VII. *Compute the new stress (i.e., repeat Step III).* Applying formula (4) to the distance values of Table 10 and the disparity values of Table 11 yields.

$$S_1 = \sqrt{\frac{2.38}{66.03}} = 0.1897: \qquad S_2 = 1.00$$

(Compare with previous stress value of 0.243 in Step III above.)

The second iteration and those that follow begin with Step IV and repeat Steps V, VI, and VII. Results of these steps are given below. All necessary information for computing the results below is given in the description of the first iteration.

ITERATION 2

1. *Compute a new configuration:*

Point 1 on Axis 1
(3.35 − 3.26) × (3.90 − 4.46)/3.35 = −0.02
(4.39 − 3.26) × (1.26 − 4.46)/4.39 = −0.82
(2.80 − 3.26) × (1.68 − 4.46)/2.80 = 0.45

−.10 is mean move (alpha = 1.00)

Point 2 on Axis 1
(3.35 − 3.26) × (4.46 − 3.90)/3.35 = 0.02
(2.65 − 3.26) × (1.26 − 3.90)/2.65 = 0.60
(3.68 − 3.26) × (1.68 − 3.90)/3.68 = −0.25

0.09 is mean move (alpha = 1.00)

Point 3 on Axis 1
(4.39 − 3.26) × (4.46 − 1.26)/4.39 = 0.82
(2.65 − 3.26) × (3.90 − 1.26)/2.65 = −0.60
(2.67 − 3.26) × (1.68 − 1.26)/2.67 = −0.09

0.03 is mean move (alpha = 1.00)

Point 4 on Axis 1
(2.80 − 3.26) × (4.46 − 1.68)/2.80 = −0.45
(3.68 − 3.26) × (3.90 − 1.68)/3.68 = 0.25
(2.67 − 3.26) × (1.26 − 1.68)/2.67 = 0.09

−0.03 is mean move (alpha = 1.00)

Point 1 on Axis 2
(3.35 − 3.26) × (3.78 − 0.48)/3.35 = 0.09
(4.39 − 3.26) × (3.49 − 0.48)/4.39 = 0.77
(2.80 − 3.26) × (0.85 − 0.48)/2.80 = −0.06

0.20 is mean move (alpha = 1.00)

Point 2 on Axis 2

(3.35 − 3.26) × (0.48 − 3.78)/3.35 = −0.09
(2.65 − 3.26) × (3.49 − 3.78)/2.65 = 0.07
(3.68 − 3.26) × (0.85 − 3.78)/3.68 = −0.34

−0.09 is mean move (alpha = 1.00)

Point 3 on Axis 2

(4.39 − 3.26) × (0.48 − 3.48)/4.39 = −0.77
(2.65 − 3.26) × (3.78 − 3.49)/2.65 = −0.07
(2.67 − 3.26) × (0.85 − 3.49)/2.67 = 0.58

−0.07 is mean move (alpha = 1.00)

Point 4 on Axis 2

(2.80 − 3.26) × (0.48 − 0.85)/2.80 = 0.06
(3.68 − 3.26) × (3.78 − 0.85)/3.68 = 0.34
(2.67 − 3.26) × (3.49 − 0.85)/2.67 = −0.58

−0.05 is mean move (alpha = 1.00)

CONFIGURATION

	1	2
1	4.364	0.685
2	3.987	3.694
3	1.295	3.420
4	1.655	0.801

II *Compute the new distances from above configurations.*

DISTANCES

	1	2	3	4
1	0.0	3.033	4.111	2.712
2	3.033	0.0	2.707	3.716
3	4.111	2.707	0.0	2.643
4	2.712	3.716	2.643	0.0

III *Compute the new disparities.*

DISPARITIES

	1	2	3	4
1	0.0	3.033	3.178	3.178
2	3.033	0.0	3.178	3.178
3	3.178	3.178	0.0	3.178
4	3.178	3.178	3.178	0.0

IV. *Compute the new stress:*

STRESS

$$S_1 = \sqrt{\frac{1.89}{61.57}} = 0.1750; (S_2 = 0.9954)$$

(Compare with previous stress value of 0.1897.)

ITERATION 3

I. THE NEW CONFIGURATION

	1	2
1	4.306	0.835
2	4.020	3.601
3	1.333	3.385
4	1.641	0.779

II. DISTANCES

	1	2	3	4
1	0.0	2.781	3.916	2.666
2	2.781	0.0	2.695	3.692
3	3.916	2.695	0.0	2.625
4	2.666	3.692	2.625	0.0

III. DISPARITIES

	1	2	3	4
1	0.0	2.781	3.119	3.119
2	2.781	0.0	3.119	3.119
3	3.119	3.119	0.0	3.119
4	3.119	3.119	3.119	0.0

IV. STRESS

$$S_1 = \sqrt{\frac{1.59}{57.96}} = 0.1658; (S_2 = 0.9714)$$

Further iterations are possible (and desirable since they will result in lower stress values; and all necessary information is given above for the reader to continue these iterations. As a check on computation the results of two iterations along the path to the final solution are given below.

ITERATION 11

I. CONFIGURATION

	1	2
1	4.136	1.633
2	4.037	2.867
3	1.455	3.404
4	1.671	0.696

II. DISTANCES

	1	2	3	4
1	0.0	1.237	3.213	2.638
2	1.237	0.0	2.638	3.211
3	3.213	2.638	0.0	2.717
4	2.638	3.211	2.717	0.0

III. DISPARITIES

	1	2	3	4
1	0.0	1.237	2.883	2.883
2	1.237	0.0	2.883	2.883
3	2.883	2.883	0.0	2.883
4	2.883	2.883	2.883	0.0

IV. STRESS

$S_1 = 0.0916$; ($S_2 = 0.3731$)

ITERATION 20

I. CONFIGURATION

	1	2
1	4.072	2.046
2	4.039	2.449
3	1.483	3.455
4	1.706	0.649

II. DISTANCES

	1	2	3	4
1	0.0	0.404	2.947	2.747
2	0.404	0.0	2.747	2.947
3	2.947	2.747	0.0	2.815
4	2.747	2.947	2.815	0.0

III. DISPARITIES

	1	2	3	4
1	0.0	0.404	2.841	2.841
2	0.404	0.0	2.841	2.841
3	2.841	2.841	0.0	2.841
4	2.841	2.841	2.841	0.0

IV. STRESS

$S_1 = 0.0317$ ($S_2 = 0.0904$)

FINAL SOLUTION

After 30 iterations, stress was reduced to a value of 0.0085. Other results were as shown below:

CONFIGURATION

	1	2
1	4.054	2.194
2	4.046	2.301
3	1.488	3.465
4	1.713	0.640

DISTANCES

	1	2	3	4
1	0.0	0.108	2.864	2.810
2	0.108	0.0	2.810	2.864
3	2.864	2.810	0.0	2.834
4	2.810	2.864	2.834	0.0

DISPARITIES

	1	2	3	4
1	0.0	0.108	2.837	2.837
2	0.108	0.0	2.837	2.837
3	2.837	2.837	0.0	2.837
4	2.837	2.837	2.837	0.0

STRESS

$S_1 = 0.0085$; ($S_2 = 0.0216$)

4. The problem of the initial configuration. A recent improvement in multidimensional scaling algorithms has resulted from the search for a method of deriving an initial configuration that would be closer to the final solution configuration than an arbitrary or random one [62, pp. 18–20]. There are two reasons to support such a search. First, it is clear that fewer iterations of the algorithm will be required if the initial configuration is close to the solution configuration, thus, a considerable saving in computing time is achieved. Second, since the method of computing the monotonically transformed distances (disparities) is a weak monotonic transformation in that it often results in the tying of untied original similarities data, and since the goodness-of-fit criterion is a measurement of

the reduction in the squared differences between the configuration distances and the disparities, degenerate "solutions" are possible in which stress is very small but in which only the weak-monotonicity requirement has been met [43, p. 16]. (A strong monotonic requirement is one in which the monotonically transformed values increase when the original similarities increase and are tied when the similarities data are tied. This criterion is met by Guttman's rank-image transformation with the addition of his strong-monotonic requirement [18, pp. 479–482].) The different methods of determining initial configurations in the various multidimensional scaling approaches will not be reviewed here. Discussions are available elsewhere [43, p. 6/ 57, pp. 254–258; 62, pp. 18–20]. However, the advantage of beginning the series of iterations outlined in the First Example with an initial configuration closer to the final solution configuration can be illustrated by returning to the First Example and recomputing the solution adopting this approach. The first stage, in this case, is to apply the metric method of Torgerson [59, pp. 254–258] to the initial similarities to compute the initial configuration (Figure 6). This method consists of computing the two (since this is the proposed dimensionality of the solution being sought) largest eigenvectors of a matrix of scalar products computed from the initial similarities data.

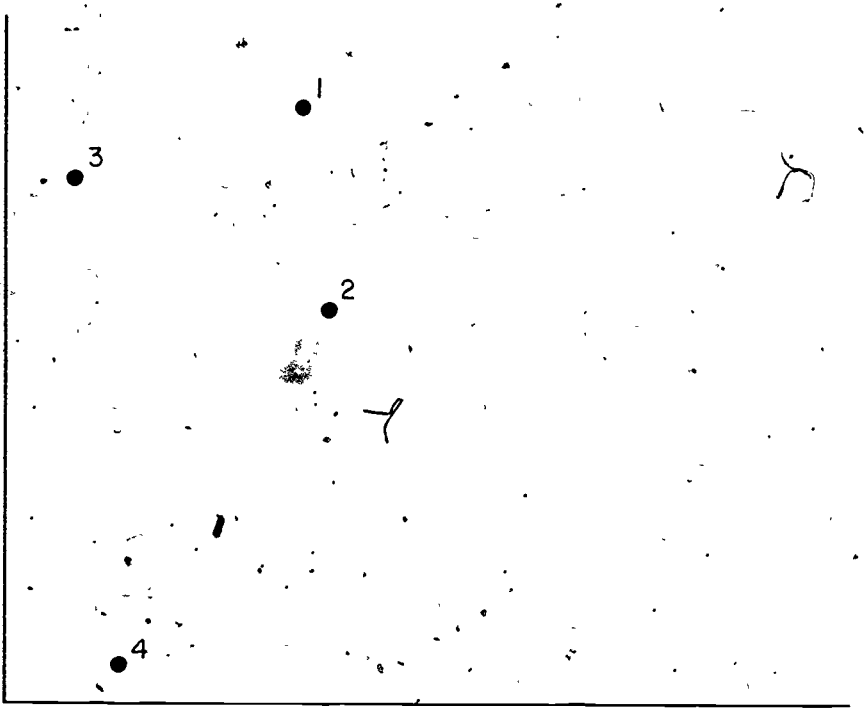


Figure 6. The Initial Configuration Derived from the Metric Approach of Torgerson on the Original Similarities

ORIGINAL SIMILARITIES (see Table 5)

	1	2	3	4
1	0.0	1.000	2.000	6.000
2	1.000	0.0	3.000	4.000
3	2.000	3.000	0.0	5.000
4	6.000	4.000	5.000	0.0

I. CONFIGURATION

	1	2
1	−0.589	−0.127
2	−0.073	−0.325
3	−0.288	0.426
4	0.951	0.025

II. DISTANCES

	1	2	3	4
1	0.0	0.553	0.629	1.548
2	0.553	0.0	0.780	1.082
3	0.629	0.780	0.0	1.302
4	1.548	1.082	1.302	0.0

III. DISPARITIES

	1	2	3	4
1	0.0	0.533	0.629	1.548
2	0.553	0.0	0.780	1.082
3	0.629	0.780	0.0	1.302
4	1.548	1.082	1.302	0.0

IV. STRESS

$S_1 = 0.0$

With the First Example, the initial configuration using Torgerson's metric method yields a better solution than the non-metric algorithm described earlier after 30 iterations. This result supports the conclusion that a better final solution will more likely be found if such a method is used to derive an initial configuration. The First Example is artificial in that it was derived from metric data, and therefore it should not be surprising that a metric method can successfully recover the original configuration from which the original similarities were derived. However, in more realistic examples, it normally is the case that substantial improvements are made to the metrically derived initial configuration by the non-metric algorithm such as the one discussed previously.

B. The Second Example

This worked example differs from the previous one in that the similarities

between points with which the scaling begins are derived from field questionnaire responses. The existence of a configuration is thus a hypothesis, and its true dimensionality is also unknown at the outset. The similarities data result from a paired-comparisons treatment of the towns chosen by a random sample of Iowa rural households for major grocery expenditures [45]. This study is described in greater detail in a later section of this paper. Specifically, the data indicate the absolute value of the difference from 0.5 of the relative frequency with which towns at one range of distance and belonging to a town-size-distance class are regarded as stimuli. Since the stimuli can be decomposed into the two component parts of town size and distance, the purpose of the scaling approach is to determine the relative trade-off between the two components. Thus, we wish to answer the following question for any and all comparisons. How much nearer or further should a town of a given size be in order to be just preferable to a second town of a given size at a given distance?

In computing these disparities, a problem resulting from ties in the original similarities is encountered. This problem did not occur in the First Example, though in practice it is a very common one. It is solved here using the "primary approach" [43, p. 9]. In this approach, an order relation within a tie is determined from the numerical order of the corresponding distances from the configuration. This is in contrast to the "secondary approach" in which the distances from the configuration are first averaged into a block corresponding to any tied values in the original similarities data. Since the primary approach may result in disparities that are different for tied values in the original data, whereas this cannot happen with the secondary approach, the former is known as the weak monotonicity approach, and the latter is known as the semi-weak monotonicity approach [43, p. 11].

The original similarities for this Second Example are shown in Table 12 below.

TABLE 12. ORIGINAL SIMILARITIES

1	0.0	0.430	0.330	0.140	0.400	0.340
2	0.430	0.0	0.500	0.410	0.500	0.470
3	0.330	0.500	0.0	0.370	0.300	0.120
4	0.140	0.410	0.370	0.0	0.500	0.390
5	0.400	0.500	0.300	0.500	0.0	0.500
6	0.340	0.470	0.120	0.390	0.500	0.0

Computing the scale in one dimension, the initial (random) configuration had the coordinates:

Point	Random Coordinates
1	0.626
2	0.940
3	0.493
4	0.713
5	0.497
6	0.907

The distances and disparities for this initial configuration are shown below.

INITIAL DISTANCES

	1	2	3	4	5	6
1		.314	.133	.087	.129	.281
2			447	.227	443	.033
3				.220	.004	.414
4					.216	.194
5						.410
6						

INITIAL DISPARITIES

	1	2	3	4	5	6
1		.199	.160	.160	.199	.199
2			.447	.199	.443	.199
3				.199	.160	.160
4					.216	.199
5						.410
6						

ITERATION 1

Point 1 on Axis 1

2 (0.31 − 0.20) × (0.94 − 0.63)/0.31 = 0.12
3 (0.13 − 0.16) × (0.49 − 0.63)/0.13 = 0.03
4 (0.09 − 0.16) × (0.71 − 0.63)/0.09 = −0.07
5 (0.13 − 0.20) × (0.50 − 0.63)/0.13 = 0.07
6 (0.28 − 0.20) × (0.91 − 0.63)/0.28 = 0.08

0.94 is mean move (alpha = 1.00)

Point 2 on Axis 1

1 (0.31 − 0.2) × (0.63 − 0.94)/0.31 = −0.12
3 (0.45 − 0.45) × (0.49 − 0.94)/0.45 = 0.0
4 (0.23 − 0.20) × (0.71 − 0.94)/0.23 = −0.03
5 (0.44 − 0.44) × (0.71 − 0.94)/0.44 = 0.0
6 (0.03 − 0.20) × (0.91 − 0.94)/0.03 = 0.17

0.00 is mean move (alpha = 1.00)

Point 3 on Axis 1

1 (0.13 − 0.16) × (0.63 − 0.49)/0.13 = −0.03
2 (0.45 − 0.45) × (0.71 − 0.49)/0.45 = 0.0
4 (0.22 − 0.20) × (0.94 − 0.49)/0.22 = 0.02
5 (0.00 − 0.16) × (0.50 − 0.49)/0.00 = −0.16
6 (0.41 − 0.16) × (0.91 − 0.49)/0.41 = 0.25

0.02 is mean move (alpha = 1.00)

Point 4 on Axis 1

1	(0.09 – 0.16) × (0.63 – 0.71)/0.09 =	0.07
2	(0.23 – 0.20) × (0.94 – 0.71)/0.23 =	0.03
3	(0.22 – 0.20) × (0.49 – 0.71)/0.22 =	−0.02
5	(0.22 – 0.22) × (0.71 – 0.71)/0.22 =	0.0
6	(0.19 – 0.20) × (0.91 – 0.71)/0.19 =	−0.00

0.01 is mean move (alpha = 1.00)

Point 5 on Axis 1

1	(0.13 – 0.20) × (0.63 – 0.50)/0.13 =	−0.07
2	(0.44 – 0.44) × (0.94 – 0.50)/0.44 =	0.0
3	(0.00 – 0.16) × (0.49 – 0.50)/0.00 =	0.16
4	(0.22 – 0.22) × (0.71 – 0.50)/0.22 =	0.0
6	(0.41 – 0.41) × (0.91 – 0.50)/0.41 =	0.0

0.01 is mean move (alpha = 1.00)

Point 6 on Axis 1

1	(0.28 – 0.20) × (0.63 – 0.91)/0.28 =	−0.08
2	(0.03 – 0.20) × (0.94 – 0.91)/0.03 =	−0.17
3	(0.41 – 0.16) × (0.49 – 0.91)/0.41 =	−0.25
4	(0.19 – 0.20) × (0.71 – 0.91)/0.19 =	0.00
5	(0.41 – 0.41) × (0.50 – 0.91)/0.41 =	0.0

I. CONFIGURATION

	1
1	0.663
2	0.944
3	0.508
4	0.725
5	0.511
6	0.824

II. DISTANCES

	1	2	3	4	5	6
1	0.0	0.281	0.155	0.063	0.151	0.161
2	0.281	0.0	0.436	0.218	0.432	0.120
3	0.155	0.436	0.0	0.217	0.003	0.316
4	0.063	0.218	0.217	0.0	0.214	0.099
5	0.151	0.432	0.003	0.214	0.0	0.313
6	0.161	0.120	0.316	0.099	0.313	0.0

III. DISPARITIES

	1	2	3	4	5	6
1	0.0	0.026	0.155	0.127	0.157	0.157
2	0.206	0.0	0.349	0.206	0.349	0.206
3	0.155	0.349	0.0	0.157	0.127	0.127
4	0.127	0.206	0.157	0.0	0.349	0.157
5	0.157	0.349	0.127	0.349	0.0	0.349
6	0.157	0.206	0.127	0.157	0.349	0.0

IV. STRESS

$$S_1 = \sqrt{\frac{.11}{.89}} = 0.3497$$

ITERATION 2

I. CONFIGURATION

	1
1	0.666
2	0.915
3	0.544
4	0.741
5	0.517
6	0.793

II. DISTANCES

	1	2	3	4	5	6
1	0.0	0.249	0.123	0.075	0.150	0.127
2	0.249	0.0	0.372	0.174	0 399	0.122
3	0.123	0.372	0.0	0.197	0.027	0.250
4	0.075	0.174	0.197	0.0	0.224	0.052
5	0.150	0.399	0.027	0.224	0.0	0.277
6	0.127	0.122	0.250	0.052	0.277	0.0

III. DISPARITIES

	1	2	3	4	5	6
1	0.0	0.186	0.123	0.117	0.150	0.126
2	0.186	0.0	0.318	0.174	0.318	0.186
3	0.123	0.318	0.0	0.126	0.117	0.117
4	0.117	0.174	0.126	0.0	0.318	0.126
5	0.150	0.318	0.117	0.318	0.0	0.318
6	0.126	0.186	0.117	0.126	0.318	0.0

IV. STRESS

$S_1 = 0.3081$

ITERATION 5

I. CONFIGURATION

	1
1	0.681
2	0.873
3	0.617
4	0.733
5	0.508
6	0.764

II. DISTANCES

	1	2	3	4	5	6
1	0.0	0.192	0.064	0.053	0.172	0.083
2	0.192	0.0	0.256	0.140	0.365	0.109
3	0.064	0.256	0.0	0.117	0.108	0.147
4	0.053	0.140	0.117	0.0	0.225	0.031
5	0.172	0.365	0.108	0.225	0.0	0.256
6	0.083	0.109	0.147	0.031	0.256	0.0

III. DISPARITIES

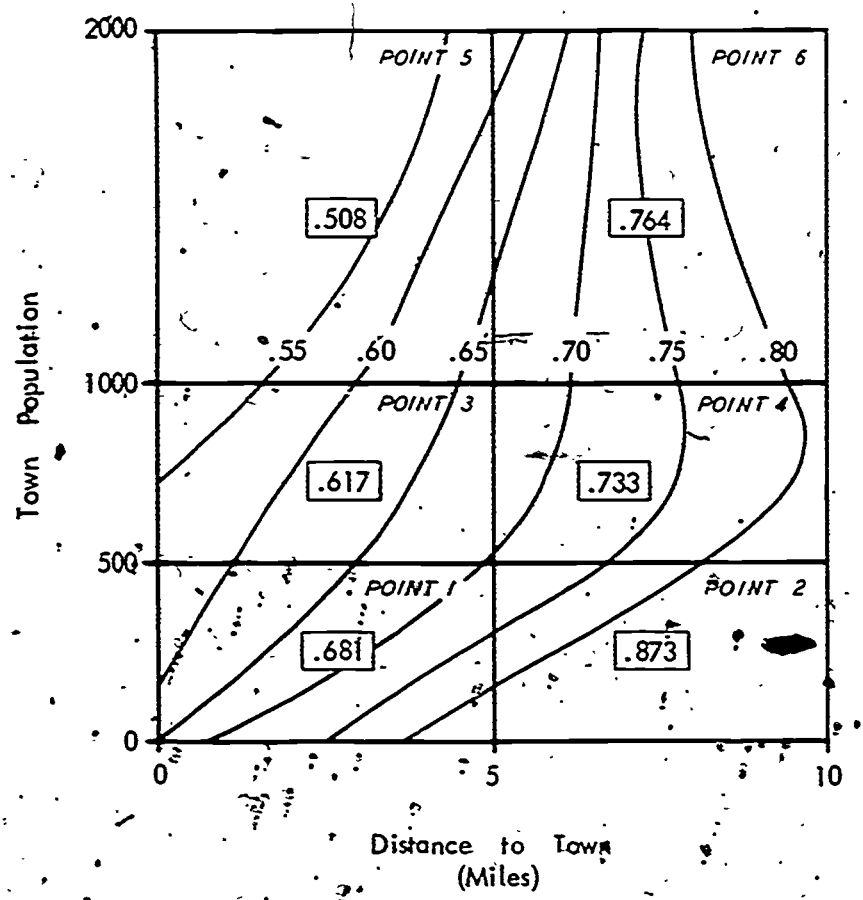
	1	2	3	4	5	6
1	0.0	0.153	0.086	0.068	0.153	0.086
2	0.153	0.0	0.275	0.153	0.275	0.153
3	0.086	0.275	0.0	0.086	0.086	0.086
4	0.086	0.153	0.086	0.0	0.275	0.086
5	0.153	0.275	0.086	0.275	0.0	0.275
6	0.086	0.153	0.086	0.086	0.275	0.0

IV. STRESS

$S_1 = .230$

After five iterations, the value of stress was .230; no further reduction in this value took place during the next 25 iterations. Substantive interpretation of the resulting configuration is facilitated by Figure 7 where the scale values are isoplethed for graphical display of the tradeoffs between the two conflicting stimuli of distance and town size of consumer movement. From this figure it is apparent that for the purposes of making grocery expenditures, towns of less than 1,500 population owe to proximity much of the patronage which they receive from the rural population. In general, the Iowa rural population substitutes stores in small towns

at close distances for stores in larger towns at greater distances. The advantage of the results of the scaling procedure is that, for the first time, we are able to determine exactly the circumstances in which one town will be more favorable than another for any and all rural locations.



Legend: Point numbers reference the stimuli numbers in the tables in which scale values are computed.

Numbers in boxes are scale values after the 5th iteration.

Figure 7. Isopleth of Scale Values After the Fifth Iteration Interpolated Preference Structure

III. GEOGRAPHICAL APPLICATIONS OF MULTIDIMENSIONAL SCALING ANALYSIS

The variety of actual and potential applications of MDS is large and ranges over many subsets of the field of geography. Perhaps the simplest way to arrange examples is to divide them into those that use simple space configuration (i.e. configurations of only the stimulus objects) and joint space configuration (i.e. those which map the configurations of both individuals and stimulus objects).

A. Simple Space Outputs

We have constantly stressed that MDS programs can accept both metric and non-metric input. it seems reasonable therefore to give specific examples of each type of research problem.

1. Map transformations. Probably one of the most quoted examples of metric input to a MDS program with simple space output, is Greenberg's "Roadmap" problem [17]. The basic data here were the inter-city road mileages between all pairs of 15 cities in the United States, that is, 105 or $\frac{n(n-1)}{2}$ inter-city distances. The data were interpreted as similarities data by ranking the distances with the shortest road distance represented by rank 1 and the largest distance by rank 105. The output consisted of a configuration of points in two dimensions (north-south and east-west). Discrepancies between predicted configurations—obtained from a non-metric multidimensional scaling program – and actual configurations were for the most part small and could be accounted for by the simple fact that road distances are frequently not the shortest distances between places but reflect detours around natural and man-made barriers. In other words, locations outputted by the program represent "true" locations if all road connections between the pairs of cities were straight lines. The solution here represents a type of map transformation similar to that which would be achieved if the places were located on an elastic map and joined by lines representing actual roadways, and if the elastic were then stretched in each direction until all road lines were straightened out.

Note that in this example where metric input is used, the first step in the Kruskal algorithm is to convert the metric data to non-metric (ranked) form. Thus the final configuration of points is obtained from non-metric information. This is true of all analyses in non-metric multidimensional scaling. However, where the researcher can have confidence in the metric information he begins with, it is often advantageous to resort to a metric multidimensional scaling. Tobler [54] has pointed out that the problem of constructing map projections is one in which the final metric is usually known to be Euclidean and the number of dimensions known to be two. Accordingly, he argues that empirical map projections might be designed to produce, from a matrix of empirical disfances between points, the "best" configuration of the points in two dimensions such that the sum of the squares of the difference between the original distances and the resulting configuration (map)

distances are minimized. In other words, instead of beginning with a definition of properties in the abstract that must be preserved, his procedure would be to begin with observed distances between points in the space and design a projection to best replicate these measurements. After he had computed the 2,080 spherical distances between a set of 65 regularly spaced latitude and longitude intersections covering the United States, Tobler used Torgerson's metric multidimensional scaling [57, pp. 255-258] to derive the plane map projection coordinates for the 65 points. Comparing the distances measured from this recovered configuration with the original spherical distances, he showed that the distortion values were generally less than two per cent. These distortion values, he concluded, compare favorably with those on Albers' and Lambert's conical projections with two standard parallels.

In a second example of the use of scaling procedures to construct maps, Tobler, Mielke, and Detwyler scaled the geobotanical distances between New Zealand and some neighboring islands using inter-island distances inferred from a model of the diffusion of plant species [55]: Basically, the authors attempted to examine the degree to which floristic similarities between New Zealand and its neighboring islands (of which there are eleven in all) could be explained by two geographic factors. (a) the relative position of the islands and (b) the size of the islands. The critical question was what proportion of the commonality of plant species could be explained fully by these two factors? The model they constructed attempted to answer this question by using floristics relations to define geobotanical distances which were then compared statistically with the islands' relative locations on the surface of the earth. In other words, they attempted to identify quantitatively the floristic relations of localities. They then used these relations (expressed by the number of species common to pairs of islands) together with island size and assumed interaction between islands to draw their geobotanical map. The distances between islands on the map were then compared with great circle distances to give some measure of the model's worth. Both the geobotanical distances and the great circle distances were inputted into a multidimensional scaling program (Guttman-Lingoes SSA-1) and outputted in a two-dimensional Euclidean space. In this way, the authors obtained an empirical map projection which preserves positional relations in the least squares sense more accurately than any other possible map projection. The actual fit of the maps to the distances was approximately 98%. In specifying the output configuration in these terms, the authors simplified the interpretation of the dimensions of each configuration for they represented merely the north, south, east, and west dimensions common to any other map projection.

D. G. Kendall [23] has extended the concept of recovering spatial coordinates to that of recovering temporal order. He has shown how multidimensional scaling can be used to recover the temporal sequencing of a set of data in which it is hypothesized that events occurring at a point in time contain information on occurrences that are known to have been "in vogue" over periods of time. In his example, the objects are tombs and the occurrences relate to the presence or absence of varieties of objects. The input square similarities matrix (of tombs) is the number of varieties common to the i^{th} and j^{th} tombs. Kendall shows how with

both this data as well as with hypothetical data, the recovered configuration in two dimensions is a horseshoe form. Ordering the tombs from the order in which they appear as one moves around the horseshoe leads to a seriation of the tombs that is strikingly similar to the serial positions of the tombs derived by classical archaeological seriation principles. This raises the intriguing possibility for historical geographers of designing research studies such that one might simultaneously hope to recover a spatial-temporal series. Wilkinson [59] has followed up Kendall's studies of seriation in archaeology using MDS by analyzing abundance matrices (i.e. matrices in which the values are non-negative and are weakly unimodal in each column) for the occurrence of Hamiltonian circuits. Hamiltonian circuits are defined as re-entry paths passing through each of m vertices in a linear graph only once. In the examples Kendall studied, "graves vs. artefacts" matrices were compiled. Wilkinson argued that the length of circuit in such matrices represents the sum of changes of fashions in neighboring graves and allows for the development of a measure for the overall rate of change during a given period. The minimal Hamiltonian circuit gives the minimum period for a series of changes to take place. For a given set of similarities data, finding this circuit involves ordering the terms column-wise, forcing a solution in two dimensions, preserving the order of interpoint distances "as well as possible" [59, p. 14], and then observing if there is a clear Hamiltonian arc. The presence of such an arc represents a check on the seriation in a configuration produced by MDS methods.

2. **Preferences for politicians.** The uses of MDS in political science and political geography afford another illustration of simple space output, but this time nonmetric input is used. Consider a situation where subjects are asked to state their preferences for political candidates. In an experiment conducted on 1,000 members of the Consumer Mail Panels and a selection of sociology students, Johnson [21] asked for positions on 35 political statements and selection of two from a list of 14 prominent political figures. It was suggested that rankings obtained from such preference data could be converted to their implied paired comparisons and, by adding over sample members, paired comparison proportions could be obtained which were then summarized by Bradley-Terry scale values. These scale values are a collection of numbers which sum to one, having the characteristic that the proportion of individuals preferring stimulus "A" from among any collection of stimuli is estimated by the scale value of "A" divided by the scale values for the whole collection of stimuli.

The resulting similarities data could then be inputted into a standard MDS program, a configuration similar to that produced by multiple discriminant analysis, as shown in Figure 8, would then be obtained. The naming of the dimensions in Figure 8 was based on an examination of the responses which subjects made on the extensive questionnaire concerning policy problems. Figure 9 shows the other dimensions which were inherent in the initial questioning and the position of each political figure in relation to those dimensions. Apparently, the configuration could be recovered in two dimensions with the "liberal/conservative" and "government

involvement" axis appearing to summarize best the whole range of possible axes. From the final configuration, of course, metric distances between pairs of points can be calculated and statements made as to the perceived "distance apart" of individuals on each of the dimensions.

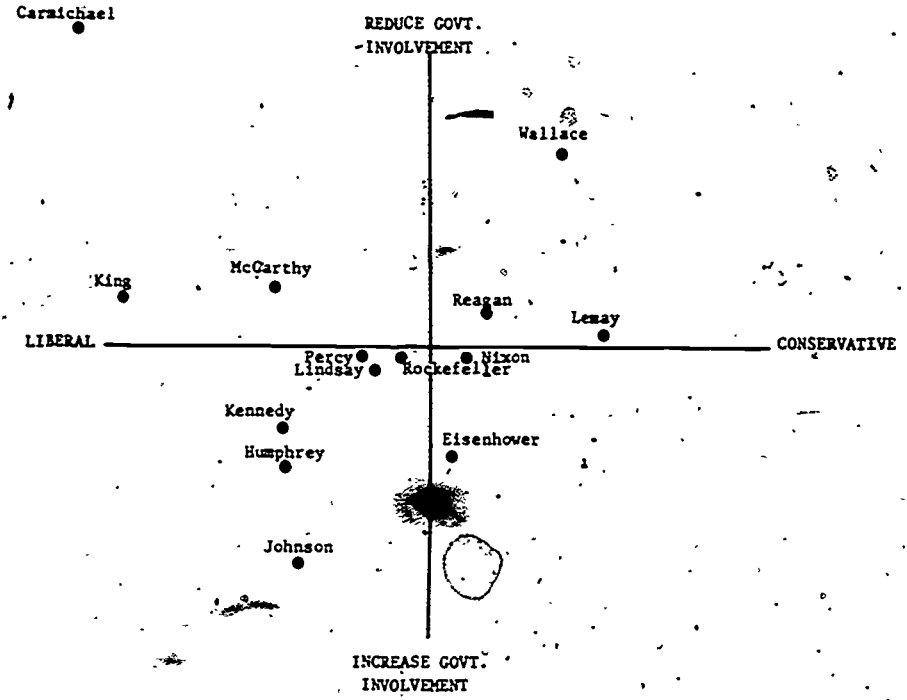


Figure 8. Positioning of Political Figures in Terms of the Two Key Dimensions (After Johnson, 1969)

3. Perceptual studies. Perceptual distance is another topic of study that appears eminently suited to analysis by MDS techniques (with simple space output). Two specific examples are the Golledge, Briggs, Demko study of intra-urban distances [14], and Whipple and Niedell's study of black and white perceptions of stores in Buffalo (N.Y.), [58].

In the study of intra-urban distances, subjects (all located at one point) estimated distances for the $\frac{n(n-1)}{2}$ pairs of locations selected for the study. The estimates obtained in this way were interpreted as dissimilarities data, and the Kruskal IV MDS program was used to produce a configuration based on interpoint distances. Figure 10 shows the configuration of points derived from the subjects' estimates of location. Since the scale of the analysis was quite small, considerable accuracy of distance estimates was obtained by some sample members. Breaking the whole sample down into two groups based on length of residence, and the distances into subsets "toward the CBD" and "away from CBD" revealed some interesting

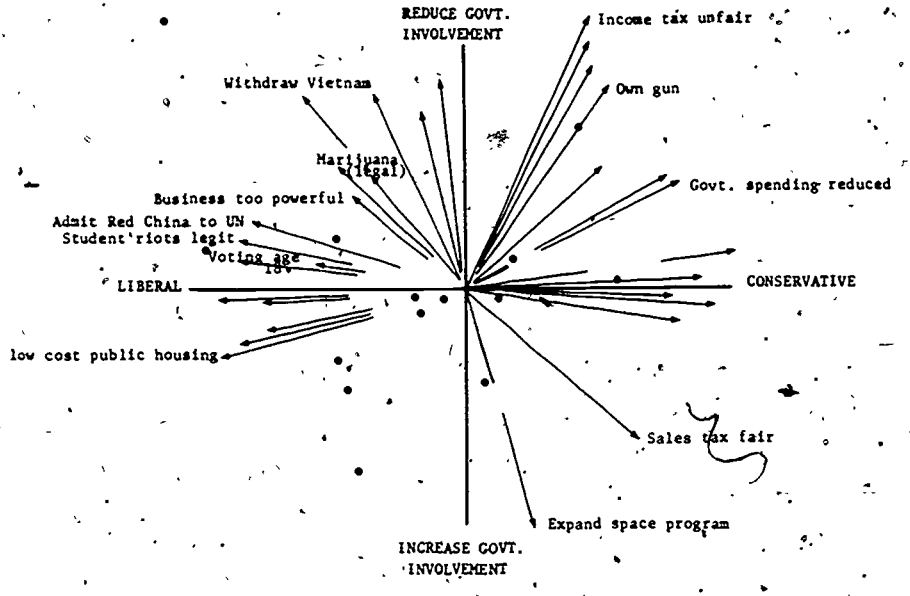


Figure 9. Positioning of Political Figures in Terms of the Two Key Dimensions (After Johnson, 1969)

trends (Figure 11). For example, the newer of the two groups was typified by much poorer estimation of distances, indicating that lack of familiarity with locations distorts their estimates of distance and can obviously influence behavior with respect to these locations. Members of the second group showed improved ability to locate all features accurately. The variance between the two groups was seen as an indication of different familiarity levels with the city, with corresponding differences in the rates of forming travel habits, and differences in the choices of orientation nodes about which mental images of the urban area were built up.

The other significant feature derived from this analysis was the tendency to exaggerate distances toward the CBD. This in turn suggests that increasing congestion (and travel time) tend to increase the perceived distances between places, and that the denser packing of land uses around the CBD makes distances appear longer and individual places harder to locate precisely.

Conclusions drawn from this study were that interpoint distances which are over-estimated probably reduce the likelihood of interaction between points. Suggestions were made as to the likely effect of distorted distance perceptions on things such as places chosen to shop, recreate, and establish residence. It seems reasonable to assume also that further studies of this type will throw considerable light on the relationships between perceptual accuracy and movement, and on the effects that changing configuration (resulting from information changes) have on urban spatial behavior.

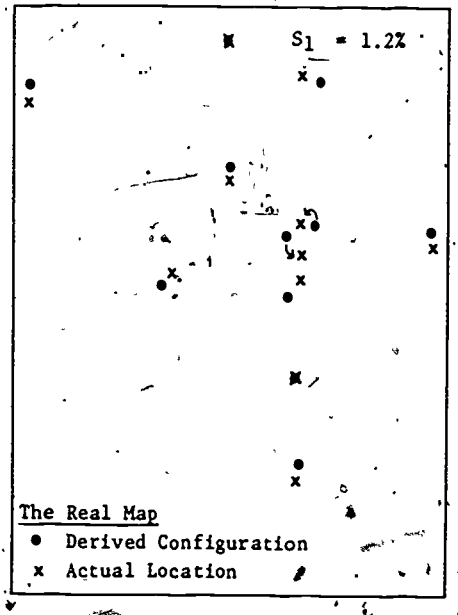


Figure 10. Actual and Calculated Locations of Sample Points. The actual map is compared to the configuration derived from dissimilarity measures consisting of actual distances between points on the real map. (After Golledge, Briggs, and Demko, 1969)

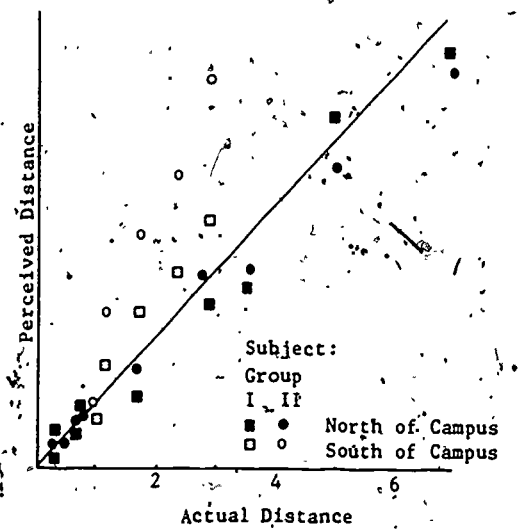


Figure 11 Perceived and Actual Distance Relationships. Perceived distances to places along a major north-south arterial passing through the campus. (After Golledge, Briggs, and Demko, 1969)

Whipple and Niedell's study of black and white perceptions of various stores in Buffalo (N.Y.) provides the geographer with an interesting framework for the analysis of perceptual distances. The authors initially used a semantic differential scale to obtain a ranking of ten department stores (based on "favorableness") by black and white respondents. Frequencies of visit to each store were also collected and again stores were rank-ordered for the whole sample. Individual semantic differential scores were transformed to a "distance" measure using the following formula:

$$\left| d_{ij} \right| = \sum_{a=1}^{n} \left| x_{ia} - x_{ja} \right|$$

where $|d_{ij}|$ = absolute distance between a pair of stores,
x_{ia} = semantic score of word pair *a* for store (i),
and x_{ja} = semantic score of word pair *a* for store (j).

The result was a distance matrix of perceived similarities for the $\frac{n(n-1)}{2}$ pairs of stores; this constituted the basic input to the MDS algorithm. The results of this analysis were most revealing:

a) stores that cluster together in the final configuration are more "competitive" than those that are far apart;
b) black and white perceptions of the favorability of stores varied somewhat, but overall the perceptions were quite similar;
c) both samples did not necessarily shop at the places with the most favorable image;
d) further study based on social and economic class differences showed very little variation in the perceived favorability of stores.

While the study was undertaken in an integrated neighborhood and would therefore not exactly mirror variations in perceptions resulting from locational segregation, the methods and results indicate that perceptual distances between competitors may be a useful variable in consumer behavior studies. A modification of this to find the perceptual distances of stores *from* consumers (using joint space procedures) would probably be even more useful to the geographer.

A further example of the use of multidimensional scaling in the simple space sense is provided by Schwind [49]. Schwind's interest is in the migration distances between states in the United States. The basic input data are dyadic in nature, and the algorithm used is the Guttman-Lingoes smallest space analysis program. Schwind generates interregional dissimilarities data on the basis of migrant moves and produces a configuration of the states of the United States in which proximity relations are transformed somewhat on the bases of the migration inputs. Schwind analyzed both dyadic *streams* of movement (i.e. the net migration rates between every pair of states) and dyadic *rates* of movement. Here the ratio of net dyadic to gross dyadic migration is defined as: $m_{ij} = \frac{M_{ij} - M_{ji}}{M_{ij} + M_{ji}}$ where: m_{ij} is the dyadic ratio,

and M_{ij} , M_{ji} represent the directionally oriented flows between any two states i and j.

In this case again, only the lower half of the migration matrix was read in as data, and the solution was obtained under conditions of semi-strong monotonicity and weak monotonicity. Results were produced for each dimension up to ten. Results were interpreted both as dissimilarities and similarities data. Schwind noted that solutions based on similarities data had lower stress values than those based on dissimilarities data. He also inferred that a three-dimensional solution was most appropriate (by examining the Shepard diagram of the result). Interestingly, Schwind found dyadic net migration *streams* to be negatively associated with geographic distance. He argued on the basis of these results that it is justifiable to treat dyadic net streams as similarity data and dyadic net rates as dissimilarity data. The output from his study included: matrices of derived interstate migration distances in a space of specified dimensionality; the geometric coordinates of each state on each recovered dimension; the distance of each point or state to the origin of the r-dimension space; and graphic presentation of the position of states in the migration space (see Figures 12 and 13).

Removed to conform with copyright laws.

Figure 12. Migration Spaces: Similarities Data (Source: Schwind, 1971)

Schwind's paper is an interesting one, for it emphasizes one of the major problems involved in multidimensional scaling analysis—the interpretation of dimensions. Upon examining the geometric coordinates of states on the recovered dimensions, Schwind argued that the coordinates did not seem to reflect any obvious scaling of states on the bases of income, urbanization, climate, and so on. He did not, however, attempt to correlate the coordinate values with any scale values

derived for the appropriate explanatory variables. He did, on the other hand, attempt to interpret the matrices of derived distances of states from the origin of the migration space. Again his intuitive interpretation was that "distance to origin" values seemed to suggest that states known for high rates of in-migration are close to the center of a migrant's perception space, and that states known for high rates of out-migration are far toward the periphery.

Another application of multidimensional scaling analysis, this time using a Kruskal-type algorithm, is seen in Gould's analysis of space preference measures with respect to the residential desirability of various states in the United States [15, 16]. Whereas Schwind used net migration rates between states to give him some indication of the similarity and dissimilarity of states, Gould obtained preference orderings of the states from a sample of 25 resident graduate geography students at Pennsylvania State University. In addition to obtaining these ordered data, Gould collected interval-scaled data on the relative advantages of states. Initially, Gould considered the point configuration of 51 states in a two-dimensional Euclidean space. The arrangement of points in this space was interpreted as indicating the similarity of states over the range of subjects. The stress value, incidentally, was .224 in 2 dimensions. Figures 14 and 15 indicate the point configurations with both interval and ordinal scale measures. The scaling devices produce interesting clusters of states with perceived similarities. The ordinal scale produced a more circular distribution of points and consequently one that was comparatively easier to interpret. Gould suggested that the configuration resulting from the interval

Figure 13. Migration Spaces: Dissimilarities Data (Source: Schwind, 1971)

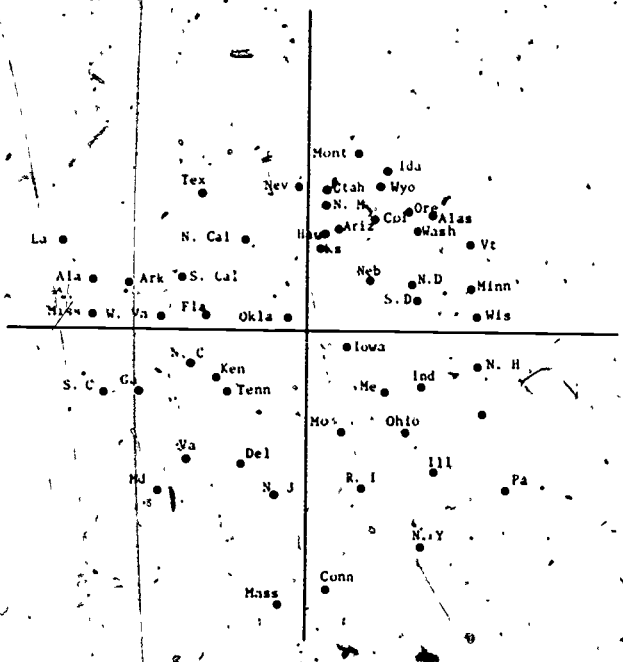


Figure 14. Point Configuration of Q-Mode Analysis With Interval Scales (After Gould, 1969)

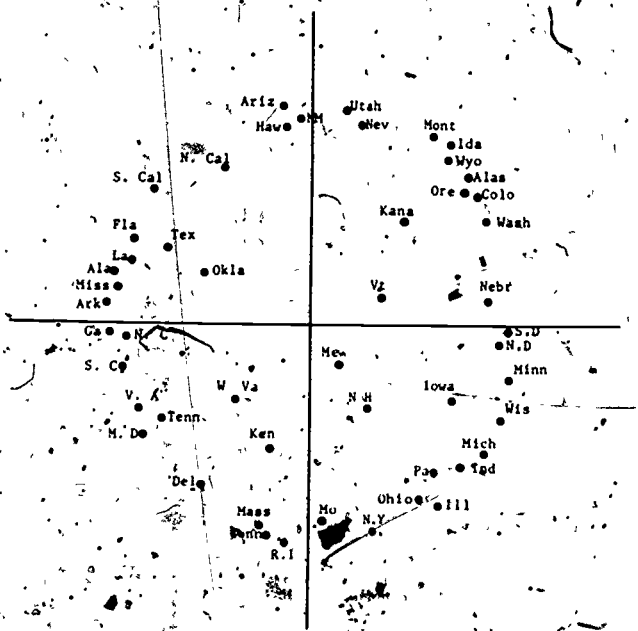


Figure 15. Point Configuration of Q-Mode Analysis With Ordinal Scales (After Gould, 1969)

scaled data might be interpreted as a map of America after it had been transformed into "some highly distorted perceptual space." He argued that it was recognizable as a map because states that occur close together in geographic space tend to be similar when perceived in terms of residential desirability.

Gould then examined the problem of whether or not the interval scale configuration is simply the ordinal scale configuration which had been randomly disturbed. In order to examine this hypothesis, he measured the locational shifts between the interval and ordinal point configurations, collected them at a common origin (similar to the collection of migratory movements for the purpose of estimating a mean information field), and examined the angles of orientation. His conclusion was that the distribution of angles can be considered as having been drawn from a rectangular distribution. He then regressed an index of social welfare for each state against the perceptual score (scale value) of each state. His conclusions ($R^2 = 0.61$) indicated that the overall mental map of the group reflects the variation of relevant welfare measures to a high degree. Further analysis of this regression provided some interesting comments on the major residual values in terms of the underestimation of the images of certain states. Finally, by finding the configuration of individuals in the sample he was able to check to see whether or not individuals located close together in his configuration had similar output configurations for their preferences for states. Again he found a high degree of correspondence between the proximity of individuals in the configuration space and the configuration of states in the state configuration space.

One of the significant conclusions from this study was that the ordinal production of data provided a more easily interpretable and clearer configuration than did the more rigorous interval scale data.

4. Archaeological reconstructions. For one final example of simple space output, we can again return to the work of Tobler and Wineberg [56]. Using the Guttman-Lingoes smallest space analysis program Tobler produced a map of pre-Hittite cities based on information obtained from analyses of cuneiform tablets. Using some of the assumptions inherent in the well-known gravity model, Tobler hypothesized that the more frequently a place was recorded on these tablets the larger would be its size. Furthermore, the more frequently pairs of cities were mentioned together on the same tablet, the greater the link between them (either in a trade or spatial sense). Based on these frequency counts, he compiled a set of dissimilarity measures and, using them as input, reproduced a configuration of the towns themselves. Since the locations of two of the towns were known, he was able to orient his output configurations in terms of latitude and longitude and to suggest an approximate location within a radius of about 50 miles for the remaining, and hitherto unlocated, places mentioned on the tablets. The essence of this study was to reproduce an archaeological or historico-geographic map of the location of places based on information inputted from a geographic model on their proximities. Information derived from the configuration may possibly then be used to choose among a large variety of alternative locations for archaeological expeditions. Incidentally, a

similar experiment attempting to reconstruct the locations of former places was conducted by Kendall, who used standardized inter-marriage rates for eight parishes in the Otmoor district of Oxfordshire, England between 1600 and 1850 as an index of similarity. Using the Kruskal MD-SCAL program he obtained a very accurate map of Otmoor [25].

While the above examples by no means exhaust the range of uses for simple space MDS in geography, they do give some idea of the types of problems that can be examined and they refer to a variety of techniques that can be used in compiling input data.

B. Joint Space Output

The use of joint space output appears to have equally as much potential use in geography as does simple space. The essence of joint space output is that both individuals and stimuli are mapped into the same dimensional space. In this way, one can obtain the subject preference rankings and at the same time give metric meaning to the distance separating individual stimuli. Concurrently, one can see how close to an individual's "ideal" any particular stimulus comes.

1. Inter-urban migration choices. One interesting use of joint space output can be seen in the work of Demko and Briggs [8] in their attempts to operationalize the choice behavior of migrants. They argued that inter-urban migration is the outcome of a choice process in which perceptions of the favorability of each alternative destination is a significant factor influencing final choice. Using a sample of individuals in southeastern Ontario, they generated similarities data concerned with the perceptions of alternative urban places and preferential choice data concerned with preferences for these places as migrant destinations.

Each individual was assumed to perceive each city as a union of attribute values. By initially mapping the cities into a perceptual space (based on similarities criteria), each place was given locational and distance characteristics somewhat different from those it possesses in physical (objective) space. In other words, places which have similar combinations of perceived attributes would be close together in the selected r-dimensional space, even though they may be far apart in objective space.

The preference model for places is derived from a multidimensional unfolding procedure. Individuals rank order places on the basis of "utility" or some other criterion of preference, then the unfolding algorithm defines an ideal point for each individual by unfolding his preferences and plotting the location of this ideal point in the same r-dimensional space as the similarities data were plotted [7, pp. 140-180]. Each city then lies a certain distance from each individual, and a one-dimensional ordering of the relations between individuals and cities can be obtained (see Figure 16). Again additional information can be obtained from the output by clustering the various locations and interpreting which places are likely to have similar drawing powers for given migrant groups.

Figure 16, Preference Space for Elmira Residents from Demko and Briggs (1970), has been removed to conform with copyright laws. It is a graph of cities and individual ideal points.

2. Scaling space preference structures. Rushton has used scaling techniques in an attempt to recover the nature of the underlying trade-offs between the various stimuli that affect the locational choice of towns for consumer expenditures by a sample of rural households [45]. His approach is described in some detail here both because it is illustrative of the use of the method of paired comparisons in generating a proximity matrix suitable for scaling by non-metric multidimensional scaling techniques, and because the four computational steps used in deriving the proximity matrix have been incorporated into an integrated computer program [26]. Before describing the four computational stages, however, a brief rationale for the researcher's interest in deriving such a scale is presented.

If our interest lies in predicting spatial choice from a set of alternatives, then we might view observed choices as the outcome of a perceptual process whereby individuals compare perceived alternatives with an ordering function of all conceivable opportunities so as to judge the most preferred alternative. Returning to reality, it might then be argued that a sensitive treatment of the places chosen–as compared with those places present but not chosen–might lead to the recovering of the exact degree of substitution of increases in increments of one variable pertaining to the places for increases in a second variable relevant to the places. Only if such statements can be made for all available alternatives, can we expect to predict choices from unique groups of available alternatives. The analogy with indifference curves in economics and preference structures in psychology has been made elsewhere [44]. In all cases, the intent of research is to specify the function that orders

all conceivable alternatives open to the individual. Since this function pertains to the relevant stimuli, the basic problem is one of scaling the stimuli.

a. Computing the input matrix. (1) Definition of stimuli. Stimuli may be defined at the outset with a sampling or other experimental design arranged so that subjects are constrained to make choices from all alternatives, or subjects might be asked to make choices from objects that are then assigned to stimulus groups. In cases where the researcher's prior knowledge of relevant stimuli is weak, the latter design is more appropriate. Rushton defined stimuli as combinations of distance-separation between people and places, and functional complexity of places (estimated by town population sizes) (Figure 17). When one town was chosen in preference to a second, the generalization was made that the stimuli group to which the first place belonged was revealed as being preferred to the stimuli-group of the second. The stimuli groups were called "locational-types." In one analysis, 30 such types were defined.

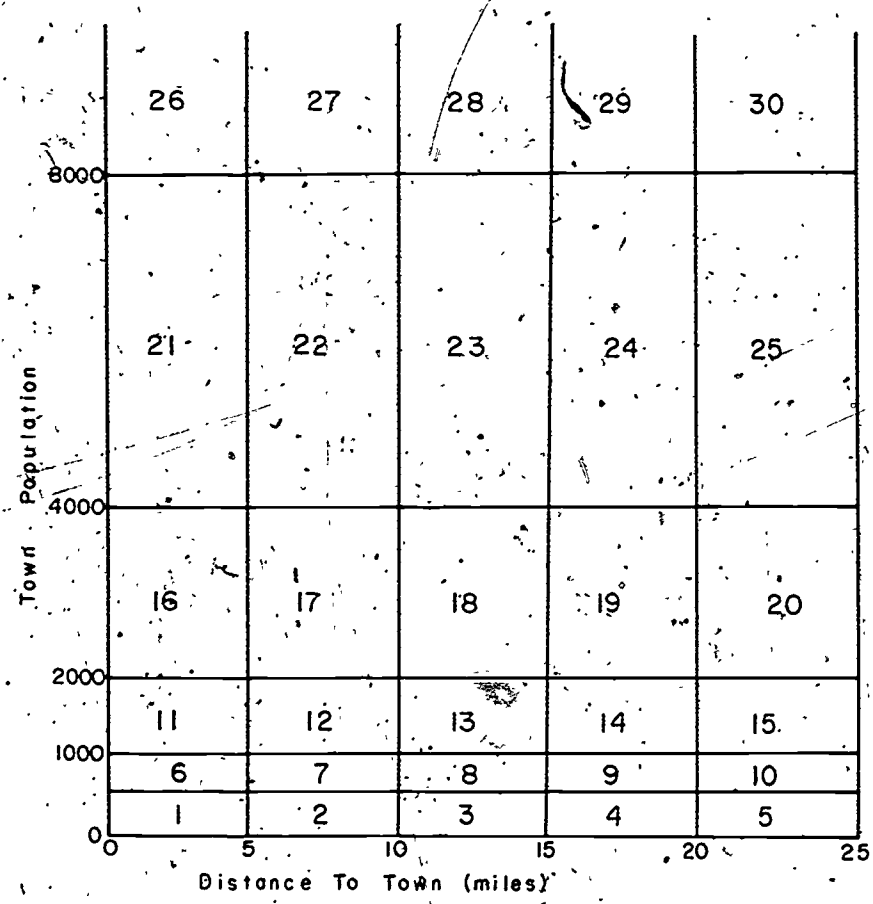


Figure 17. Definition of Locational Types (After Rushton, 1969 [45])

(2) The basic data matrix. From a random sample (603 respondents) of the rural population of Iowa in 1960, information was obtained on the places chosen for expenditures on a number of commodities. Taking the place chosen for the majority of expenditures for groceries, the following matrix was assembled.

TABLE 13. BASIC DATA MATRIX (HYPOTHETICAL)

	Locational Type							
Household ID	1	2	3	4	5	6	7 ...	30
1	0	*	*	*	0	*	1	0
2	0	0	1	0	0	*	*	*
3	1	*	*	*	0	0	0	*
4	0	*	1	*	0	0	*	0
n	*	*	*	0	0	1	*	0

Legend: Town in the indicated locational type:
1 patronized
* present, but not patronized
0 not present

b. Computing the interpoint probabilities. A measure of the extent of preference for one locational type over another is the probability that one type is chosen over another when both are present and one is chosen. This probability can be computed as a relative frequency by manipulation of the basic data matrix described above [45].

c. Computing the interpoint proximities. From this measure of the degree of preference, a measure of perceptual distance between locational types is required that, at least in an ordinal sense, will indicate closer or further perceptual distance between all pairs of types. For this purpose, we use the premise of Cattell that equally often noticed differences between stimuli are equal unless always or never noticed. Consequently, if one locational type is as frequently preferred to a second type as that type is to the first (on the occasions when both are present and one of them is chosen) the overall perceptual distance between the two types is zero. The perceived distance between any two types is given by:

$$d_{ij} = \left| p_{jpi} - 0.5 \right|.$$

That is, the perceptual distance from locational type i to type j is the absolute value of the difference from 0.5 of the conditional probability that locational type j is preferred to type i when both are present and one is chosen. It follows that $d_{ij} = d_{ji}$. A matrix of such perceptual distances is an appropriate input matrix for scaling particularly by non-metric scaling techniques, for the quality of the information is such that we are confident only of the rank ordering of the interpoint distances. However, Shepard [51] has shown that the rank ordering of all interpoint distances

in a matrix implies the metric position of the points in a space of unknown dimensionality and that, where the number of points is large (e.g., greater than 15) and the true dimensionality of the space is small, the freedom of movement in this space is most restricted if the rank order of interpoint distances between points in the metric space is to correspond with the rank order in an input matrix. Hence, he argues, that metric structures are often implicit in ordinal data.

d. Scaling the locational types. The locational types were scaled by the method of Kruskal [28], and an isopleth map of the scale values for the one-dimensional solution is included as Figure 18.

e. Significance of the recovered scale. The significance of the recovered scale is that, while "measured" from observed behavior in a spatial system of opportunities, it might explain spatial choice in a region where the density and arrangement of spatial opportunities are different. Preference scales are fundamental descriptions of behavior in that they show how all hypothetical alternatives are evaluated. Since a particular environment is a unique subset of the set of hypothetical alternatives, the preference scale may be used to evaluate this special case. It is this generality which an appropriately designed preference scale possesses that leads to its great potential in solving research problems.

f. Inter-personal comparisons of scales. Ewing [12] has compared the preference scales of locational types for different social and economic groups of the Iowa households. He has applied significance tests to the differences in probability values in the cells of an input matrix and he has also used differences in scale values for a subset of alternatives to compute the probabilities of interaction with any alternative. He found that the greatest difference in preference structures was between two groups of households, one of which had been shown to patronize the nearest available opportunity while the other was composed of households who by-passed the nearest opportunity in favor of some other. This result may be contrasted with that of Ermuth [11] who found no difference between the preference scale of one group of urban households who claimed (in a test question using the semantic differential) that the distance of a store was important in their choice, and the scale of a second group who claimed that distance was not important.

g. Temporal comparisons of scales. Several factors lead to temporal changes in such scales. Changes in how people evaluate alternatives are often induced by changes in the character of the alternatives themselves and in the transportation system that relates them to the alternatives, these may lead to changed preference scales. One study has compared the Iowa preference scale for grocery purchases in 1960 with that of 1935 [46]. Major differences found in the two scales can be summarized as an increasing tendency in 1960 for Iowa households to by-pass small towns at close distances for larger towns farther away. Such a generalization has previously been

made on the basis of less formal and less quantitative research, but the precise calibration of the change is not possible unless the scales are computed. The growth or decline of specific towns in this period will depend to a large degree on their position in relation to the two scales.

3. Preference and choice in different environments. Comparisons have been made between preference scales measured in two different environments, southwestern Michigan and Iowa [47]. The comparisons showed that, despite different distance-decay functions for the probability of choices made at different distances, the recovered preference scales are similar in the two areas.

a. Joint-space analysis of preference scales. A second approach to the question of choice in different environments is to ask whether spatial choices for like things in different areas can be regarded as different points of view from which the two groups evaluate the stimuli (locational types). Beginning with a matrix consisting of the scale rankings of the locational types for three commodities in Iowa and for the same three in Michigan, the six sets of rankings were evaluated to determine where each was positioned in the joint space with the 30 stimuli (Table 14). The locational types are defined in Figure 18. The two-dimensional joint-space solution is shown in Figure 19, and the scatter plot of configuration distances and input ranks is shown in Figure 20. The stress value which measures the goodness-of-fit is 0.052. The analysis used the TORSCA algorithm [61, p. 13]. The close proximity in this preference space of five of the six points representing the groups shows that systematic differences in viewpoints of those five groups do not exist, rather that the ordering of spatial alternatives is similar in all five cases. The sixth point, describing the viewpoint of the Michigan group choosing towns for clothing purchases, is anomalous and deserves further study. Such differences in preference structures can be attributed to one or both of two sources. They may indicate that one group evaluates similar stimuli differently, for example, the results above might mean that Iowa households have a greater propensity to patronize small, local towns whereas their Michigan counterparts might have changed their former habits and now prefer to make the longer journey to the bigger towns. Alternatively, the different preference structures might reflect the fact that similar-sized towns in Michigan and Iowa might contain different amounts and types of clothing stores, then the observed pattern of Figure 19 would be a reflection of the ambiguity present in the surrogate variable "town-size" as a measure of town content. Further research would clarify these interpretations. However, the anomalous group aside, the tight cluster of the other five groups in the perceptual space and the accuracy with which the independently computed preference structures could be recovered from this joint space is an indication of the consistency of spatial-preference structures for different trip purposes, as well as for choice of towns in different areas.

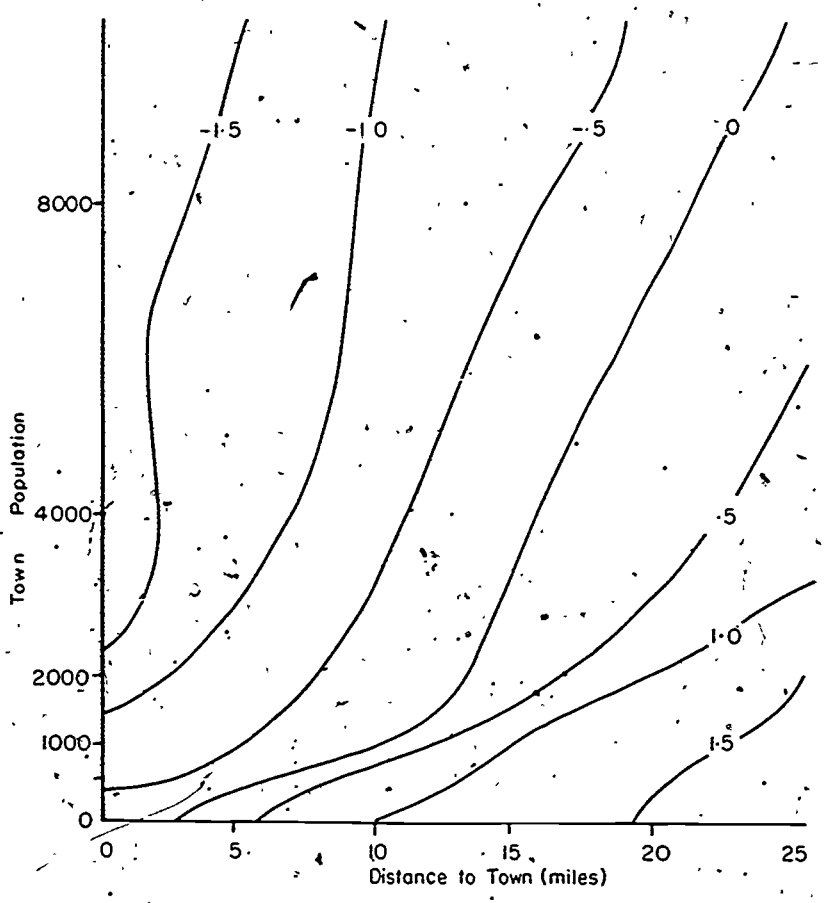


Figure 18. Space Preference Structure for Grocery Purchases. Iowa 1960 (After Rushton, 1969 [45])

An ingenious recent application of joint-space analysis was designed to shed light on the problem of interpreting scaled dimensions, referred to in Section I. In a study of individual rankings of U.S. states for residential desirability, Lieber [31] added to the m x n matrix of m states and n state viewpoints, lx m objective measures of the states on variables hypothesized to be related to residential desirability preferences. He then simultaneously scaled the m + n + l points in the same space and interpreted interpoint distances between the state viewpoints and the objective criteria as a measure of the degree to which the state viewpoint corresponded with the various objective criteria. Tables 15 and 16 show the results for the 17 most preferred states.

b. Incorporation of preference scales in diffusion models. De Temple [19] has

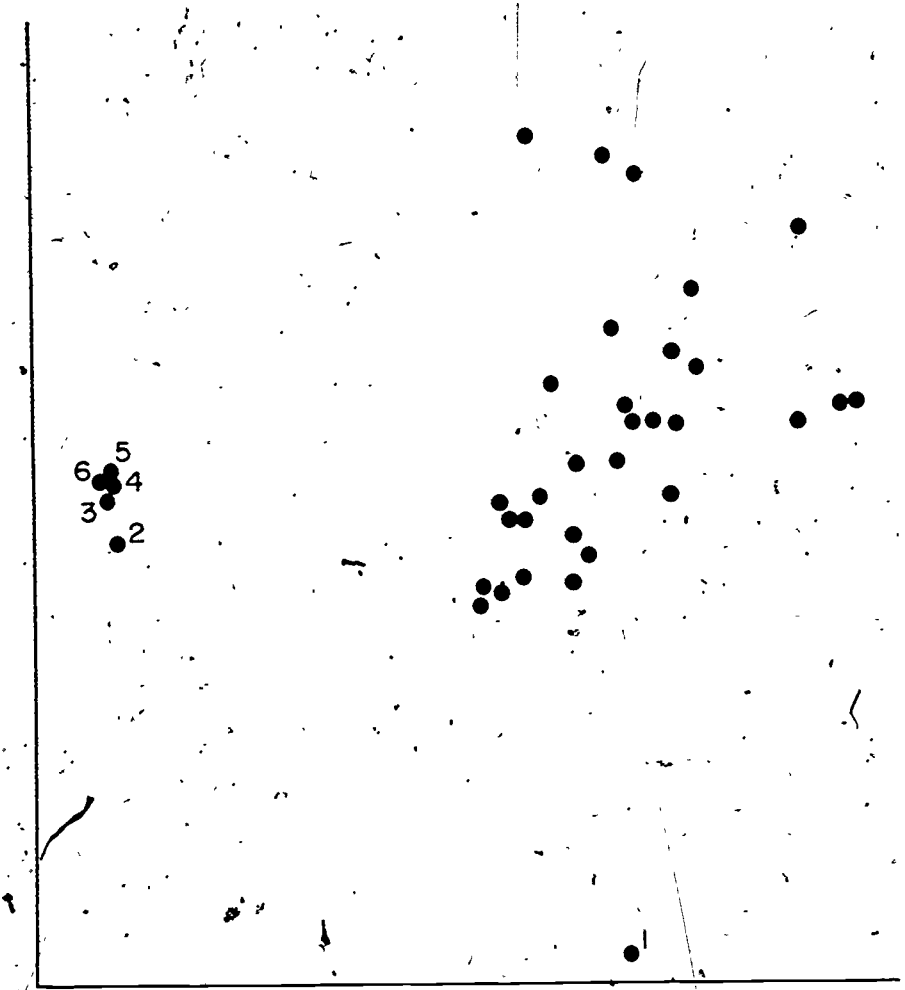


Figure 19. Joint Space for Consumer Spatial Choices in Michigan and Iowa. Legend. 30 unidentified points refer to the 30 locational types. Points numbered 1–Michigan, clothing; 2–Iowa, clothing: 3–Michigan, appliances; 4–Iowa, appliances; 5–Michigan, groceries; 6–Iowa, groceries

argued that a space-preference structure is a more appropriate predictor of spatial interaction rates than the more commonly used distance-decay functions, since, as we stated above, the preference structure is more sensitive to the unique distribution of people or places in a given context. De Temple used preference structures for towns selected for different commodity expenditures generalized for the probabilistic allocation rule of Luce [36] in order to generate probabilities of interaction with places to govern the spread of the adoption of a farming innova-

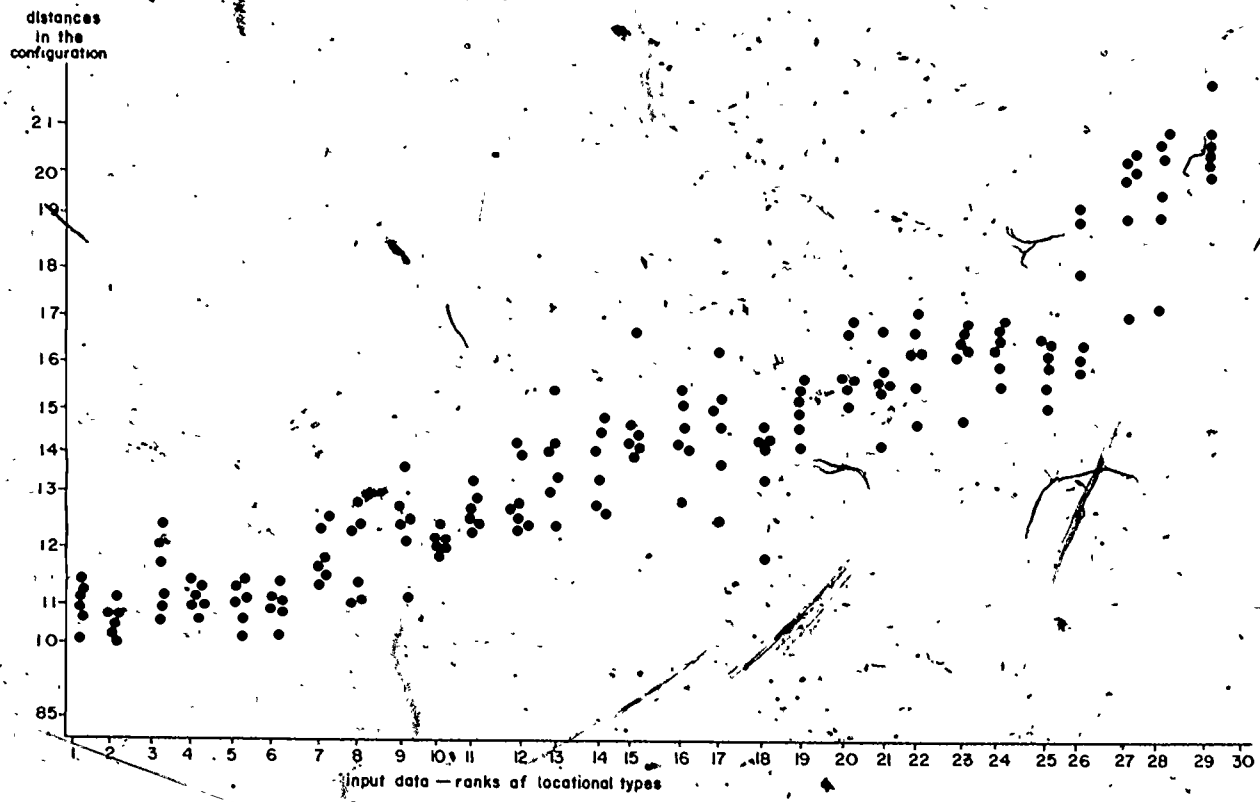


Figure 20. Scatter Plot of Configuration Distances and Input Preference Ranks

tion. Formerly, distance-interaction rates had been used for this purpose. His operationalization of the diffusion model is thus closer to the theoretical model first proposed by Hägerstrand [19]. Others experimenting with preference structures include Ewing [12] and Girt [13], who have extended the application of space-preference structures by applying Luce's choice axiom, and Briggs [5] in his study of preferences for shopping centers in Columbus, Ohio.

TABLE 14. RANK-ORDER OF THE THIRTY LOCATIONAL TYPES FOR THE SIX PREFERENCE STRUCTURES

Locational Types	Mich. Clothing	Iowa Clothing	Mich. Appliances	Iowa Appliances	Mich. Groceries	Iowa Groceries
1	16,000	21,000	13,000	16,000	12,000	15,000
2	15,000	24,000	19,000	21,000	16,000	22,000
3	28,000	27,000	21,000	24,000	20,000	25,000
4	27,000	28,000	30,000	27,000	29,000	28,000
5	25,000	29,000	29,000	30,000	28,000	30,000
6	19,000	17,000	7,000	11,000	8,000	10,000
7	23,000	18,000	15,000	18,000	14,000	17,000
8	29,000	25,000	17,000	19,000	23,000	9,000
9	26,000	22,000	28,000	23,000	22,000	24,000
10	24,000	30,000	27,000	29,000	27,000	29,000
11	10,000	8,000	4,000	5,000	4,000	5,000
12	6,000	16,000	11,000	9,000	11,000	13,000
13	18,000	19,000	16,000	25,000	17,000	19,000
14	22,000	23,000	23,000	17,000	25,000	26,000
15	20,000	26,000	26,000	28,000	30,000	27,000
16	11,000	4,000	6,000	8,000	3,000	1,000
17	7,000	11,000	12,000	12,000	7,000	11,000
18	17,000	14,000	22,000	15,000	15,000	14,000
19	30,000	15,000	25,000	26,000	24,000	21,000
20	21,000	20,000	24,000	22,000	26,000	23,000
21	9,000	3,000	9,000	4,000	5,000	4,000
22	3,000	6,000	1,000	6,000	1,000	6,000
23	13,000	9,000	14,000	3,000	13,000	12,000
24	14,000	12,000	18,000	13,000	19,000	16,000
25	12,000	13,000	20,000	20,000	21,000	20,000
26	2,000	1,000	3,000	7,000	18,000	3,000
27	1,000	2,000	2,000	2,000	2,000	2,000
28	4,000	5,000	5,000	1,000	6,000	7,000
29	8,000	7,000	10,000	10,000	10,000	8,000
30	5,000	10,000	8,000	14,000	9,000	18,000

TABLE 15. INTERPOINT DISTANCE MATRIX OF GENERAL VIEWPOINT AND OBJECTIVE CRITERION OF THE SEVENTEEN PREFERRED STATES ON THE PARTITIONED RANKINGS

State View	Alphabetic Contiguity	Urban Offerings	Warm Climate	Coastal/ Recreational	Distance	Regional Preference
Utah	0.285	1.372	1.369	1.374	1.578	1.379
Texas	1.256	0.041	0.203	0.040	0.808	0.009
Vermont	1.250	0.012	0.173	0.010	0.838	0.040
South Dakota	1.204	0.118	0.280	0.116	0.733	0.067
Tennessee	1.237	0.112	0.056	0.113	0.960	0.161
Indiana	1.356	0.272	0.434	0.271	0.583	0.222
California	1.230	0.013	0.157	0.016	0.852	0.057
Maryland	1.444	0.481	0.643	0.479	0.375	0.431
Florida	1.242	0.020	0.181	0.019	0.829	0.032

TABLE 16. FREQUENCY OF CRITERION IMPORTANCE ON THE STATE VIEWPOINTS OF PREFERRED STATES

Criterion	Most Important Factor	Second Most Important Factor
Alphabetic Contiguity	1	0
Urban Offerings	1.5	3.5
Coastal/Recreational	1.5	3
Warm Climate	1	0
Distance	1	0
Regional Preference	3	1.5
None	—	1
Total	9	9

4. Incorporation of preference scales in central-place theory. Since central-place theory describes the theoretical location of business clusters (settlements) resulting from the mutually adaptive behavior patterns of entrepreneurs and consumers, scales that describe how one of these groups responds to actions of the other group ought to be the fundamental postulates of the theory. But the scales that were used in the classical statements were so simple that they were not commonly recognized as scales. Thus, from the consumer's point of view, the postulate of Christaller that the consumer would patronize the closest place offering the required goods was essentially a scaling of relevant alternatives on the distance variable and postulating that the alternative with minimum distance scale value would be most preferred and hence patronized. Described in this way, it would seem to be a natural development for the theoretician to question the effect on the derived settlement patterns of substituting more realistic consumer preference scales for the one used by Christaller and implicitly accepted by most researchers who have "tested" the theory since his work first appeared. The question has been discussed [6], and a formal model of

central-place theory has been constructed in which the consumer preference scale has been varied [48]. The results were that, accepting all the other postulates, constraints, and environmental assumptions of Christaller, the grouping of unique bundles of goods in a central-place hierarchy was no longer a derived property of the theoretical system. Such findings support the thesis that preference scales are an important input to location theories for mutually adaptive behavior patterns which are a hallmark of all location theories.

Further evidence of the usefulness of such preferences in substantive research is found in other work unpublished at The University of Iowa. Mr. T. Bell has used trend surface equations of the scale values in a computer program that estimates tributary populations for market centers. From a close checkerboard sample of rural locations, the algorithm evaluates surrounding towns and allocates the area surrounding the sample point to the most preferred market center, using the equation for the preference function for the activity in question. Bell is currently comparing the relative sizes of tributary populations surrounding centers which have lost certain activities in the past decade with centers that have supported the activity or have added it to their business structure during that period.

IV. CONCLUDING STATEMENTS

The overall strategy adopted in this monograph has been to present MDS as a useful tool for geographers. To do this, we felt that it was necessary to explain the component parts of the method; to mention the types of metric and non-metric problems to which it had been applied, to discuss the nature of data required for input, and to examine the mathematical structure of the technique. It behooves us now to comment on some of the problems involved in using the technique, and to elaborate on the areas of potential use of MDS analysis in the discipline.

A. Some Problems of Using MDS

The problem of identifying dimensions has previously been mentioned. Perhaps the most serious problem relates to the "stopping" subroutines which determine the dimensionality and fit of the output. This is the global minimum problem. Calculation of stress values at each iteration help to determine the goodness-of-fit between a derived configuration and an actual data set. "Satisfactory" stress values can sometimes be obtained when "local minima" are reached; continued iteration may at first produce an increase in stress and then a decrease which results in values considerably lower than those obtained at the local minimum. Resulting configurations and interpoint distances may also change considerably in this process. Most MDS algorithms check for local minima by changing step sizes and continuing the iterations a given number of times. For the most part these procedures satisfy a researcher. An alternative method of handling this problem is to rerun the iterative sequence beginning with different starting configurations and checking to see if approximately the same configuration is obtained on each run. However, there is no sure way of determining if a global minimum has been reached, consequently, the technique should not at this stage be used to define such minima.

B. Potential Uses of MDS in Geography

One of the potential uses of MDS which is as yet largely unexplored in geography is that of obtaining scale values for input into other analytical algorithms. Perhaps the most obvious of these uses is the application of cluster techniques on MDS output configurations. Configurations can be analyzed in this manner either in simple or joint spaces. In the latter case, either individual or stimulus configurations can be examined.

While geographers are frequently interested in the pattern of output configurations, other disciplines are more interested in the scale values or the interpoint distances derived from MDS analyses. For example, it may be possible to use interpoint distances as input into decision models in an attempt to predict choices. In joint space analyses, distances between individuals and stimulus points become the data inputs into such models. Although no studies exist as yet which combine characteristics of clusters (such as mean areal centers) and joint space distances, it

seems that they are eminently feasible for determining likelihoods of interaction between groups of individuals and specific stimulus points.

Our summaries of the work of Tobler, Kendall, and Wilkinson illustrate one of the more creative uses of MDS analysis – that of filling in "missing data" or speculating about the locations of phenomena. Archaeologists and historical geographers appear therefore to have at their disposal an interesting and innovative method of analysis for their disciplines.

Whether joint space or simple space procedures are used, MDS appears to have a multitude of potential uses in geography. In addition to uses such as those above, we suggest the following areas of potential research:

1) to stratify populations according to specific social, economic, ethnic, or other characteristics in order to define precisely how variations in these phenomena influence perceptions of locations and attributes of things, and to determine the contribution of such stratifications to variations in preference rankings of phenomena by subjects.
2) to determine how distance or locational characteristics of configurations change with time; this partly indicates the role of learning and information gathering in obtaining perceptual accuracy. The results can then be used to help explain a variety of spatial behaviors including journey-to-work, consumer behavior, and residential site selection.
3) to assist in translating non-metric data to metric distance measures for purposes such as distinguishing clusters of like and unlike phenomena.
4) as a potential aid in policy making by determining the "perceptual gaps" that exist in groups of phenomena: for example, finding the locational and perceptual properties of stores that should be developed to serve minority and other populations.
5) to experiment with the notion of perceptual distance, to translate it into metric terms, and perhaps revise our widespread use of just two-dimensional Euclidean distances in explanatory models of spatial behavior.
6) to examine the functional relationships between perceived and "actual" distances in order to determine the range over which these distances agree, the nature of the relationship (i.e., whether linear or non-linear), the rate at which divergence occurs as distance increases, and other facts concerning the two types of distance.
7) to examine the nature of terms such as "proximity" and "closeness" in order to establish meaningful uses of the terms in spatial analysis.

The research questions proposed above have existed in a confused sense in geography for many years, the mathematical developments reviewed in this monograph aid in their clarification and formalization, providing concepts, terms, theorems, and empirical findings that stimulate the researcher's imagination. Since each question can be related to research that has been undertaken by geographers over the past few years, it would be prudent to note that in each case the question raised may ultimately be resolved by other forms of data analysis. However, experience to date has indicated the fruitfulness of MDS approaches in the analysis of

these questions. The fact that they have been raised so recently, however, illustrates how developments in basic analytical procedures stimulate researchers to ask new questions.

The development of non-metric MDS has therefore provided the geographer with concepts and techniques by which he can expect to solve some of the puzzling measurement problems that have impeded the development of behavioral geography. For example, geography has only just begun to research and to measure the form of preference structures. We can surely expect that man's adjustment to and behavior in his physical and human environment will more commonly be interpreted and researched as his reaction to a perceived set of stimuli. His evaluation of these stimuli will become a primary research problem. Decisions made in this environment will increasingly be viewed as a process by which basic preferences are linked to perceived sets of stimuli.

In addition to their role in understanding choice behavior, preference structures will increasingly be used in the normative sense of designing or controlling operating spatial systems in order to optimize subjective preference functions. Thus the central importance of basic preference studies is emphasized and the accompanying role of MDS assured.

In Section I, we mentioned the existence of several MDS algorithms. Each of these has been programmed and information regarding their availability can be obtained from either the Geography Computer Exchange Program at Michigan State University or from the authors themselves. A recent development not discussed in our text is the MINISSA series of scaling programs developed by Lingoes and Roskam, we have not yet examined this algorithm in detail but its structure and advantages are discussed by the authors [33].

In conclusion, we reiterate our position that although MDS is still in a developmental stage, it provides a useful and constructive methodology for examining the problems of preference and choice which are of increasing concern to researchers in geography, and we express the hope that our treatment of the problem will increase knowledge and availability of the method in the discipline.

REFERENCES

1. Atkinson, R. C., G. H. Bower, and E. J. Crothers. *An Introduction to Mathematical Learning Theory*. New York: J. Wiley and Sons, (1965).
2. Baskin, C. W. *Cristaller's Central Places in Southern Germany*. Englewood Cliffs, N. J.: Prentice-Hall, Inc., (1966).
3. Berry, B. J. L. and W. Garrison. "The Functional Basis of Central Place Theory," *Economic Geography*, 34 (1958), 145–154.
4. Berry, B. J. L. and N. S. Ginsberg. *Atlas of Economic Development*. Chicago, Ill.: University of Chicago, Department of Geography, (1963).
5. Briggs, R. *The Scaling of Preferences for Spatial Locations. An Example Using Shopping Centers*. Columbus, Ohio: M. A. Thesis, Department of Geography, The Ohio State University, (1969).
6. Clark, W. A. V. and G. Rushton, "Models of Intra-Urban Consumer Behavior and Their Implications for Central Place Theory," *Economic Geography*, 46 (1970), 486–497.
7. Coombs, C. H. *A. Theory of Data*. New York: J. Wiley and Sons, (1964).
8. Demko, D. and R. Briggs, "An Initial Conceptualization and Operationalization of Spatial Choice Behavior. A Migration Example Using Multidimensional Unfolding." *Proceedings of the Canadian Association of Geographers*, (1970), 79–86.
9. De Temple, D. J. "A Space Preference Approach to the Determination of Individual Contact Fields in the Spatial Diffusion of Innovations," Unpublished Manuscript, Department of Geography, Indiana State University, (August 1970).
10. Edwards, A. L. *Techniques of Attitude Scale Construction*. New York. Appleton-Century-Crofts, (1957).
11. Ermuth, F. "Intra-Urban Space Preferences for Grocery Shopping" *Pre-Conference Papers*, Canadian Association of Geographers, (1970), 117–123.
12. Ewing, G. "Consumer Space-Preferences," Unpublished Ph.D. Dissertation, McMaster University, (June 1971).
13. Girt, J. L. "A Constant Utility Model of Spatial Choice," *Pre-Conference Papers*, Canadian Association of Geographers, (1971), 68–70
14. Golledge, R. G., R. Briggs, and D. Demko, "Configurations of Distance in Intra-Urban Space." *Proceedings of the Association of American Geographers*, Vol. 1 (1969), 60–65.
15. Gould, P. R. "Structuring Information on Spatial-Temporal Preferences." *Journal of Regional Science*. Vol. 7, No. 2 (Supplement), (1967).
16. Gould, P. R. "Problems of Space Preferences, Measures and Relationships." *Geographical Analysis*. Vol. 1, No. 1, (1969), 31–44.
17. Greenberg, M. G. "A Variety of Approaches to Non-Metric Multidimensional Scaling." Paper presented at 16th International Meeting of the Institute of Management Sciences. New York, (1966).
18. Guttman, L. "A General Nonmetric Technique for Finding the Smallest Coordinate Space for a Configuration of Points," *Psychometrika*, 33 (1968), 469–508.
19. Hägerstrand, T. "A Monte Carlo Approach to Diffusion," *Archives europeennes de sociologie*, 6 (1965), 43–67.
20. Isaac, P. D. "Dissimilarity Judgments and Multidimensional Scaling Configuration as Indices of Perceptual Structure," *Michigan Mathematical Psychology Program*, 3, (1968).
21. Johnson, R. "Relationships Between Product Attributes and Preferential Choice Behavior," mimeo, (1969).
22. Kendall, D. G. "Incidence Matrices, Interval Graphs and Seriation in Archaeology," *Pacific Journal of Mathematics*, Vol. 28, No. 3 (1969).
23. Kendall, D. G. "A Mathematical Approach to Seriation," *Phil. Trans. Royal Society*, London A. 269, (1970) 125–135.
24. Kendall, D. G. "Abundance Matrices and Seriation in Archaeology," *Z. Wahrscheinlichkeitstheorie verw*, Feb. [illegible] (1971) 104–112.

25 Kendall, D G. "Construction of Maps From 'Odd Bits of Information,'" *Nature*, 231 (May 21, 1971), 158–159.

26 Kern, R. and G. Rushton *REVPREF Paired Comparisons Analysis from Revealed Spatial Preference Data*. Technical Report No. 95, Computer Institute for Social Science Research, Michigan State University, (May 1969).

27 King, L. J *Statistical Analysis in Geography* Englewood Cliffs, N.J.. Prentice-Hall, Inc., (1969).

28. Kruskal, J B "Multidimensional Scaling by Optimizing Goodness-of-Fit to a Nonmetric Hypothesis," *Psychometrika*, 29 (1964), 1–27.

29. Kruskal, J. B. "Multidimensional Scaling A Numerical Method," *Psychometrika*, 29 (1964), 115–129.

30. Kruskal, J. B and J. D. Carrol. "Geometrical Models and Badness-of-Fit Functions," in P. R. Krishnaiah (ed.) *International Symposium on Multivariate Analysis, Dayton, Ohio, 1968*, New York: Academic Press, (1969), (In press).

31 Lieber, S. "A Joint Space Analysis of Preferences and Their Underlying Traits," (Iowa City, Department of Geography, Discussion Paper No. 19, July 1, 1971).

32. Lingoes, J. C. and L. Guttman. "Nonmetric Factor Analysis, A Rank Reducing Alternative to Linear Factor Analysis," *Multivariate Behavioral Research*, 1 (1967), 485–505.

33. Lingoes, J. C and E. Roskam. *A Mathematical and Empirical Study of Two MDS Algorithms*, Michigan Mathematical Psychology Program, (MMPP), 71–1 (April 1971).

34. Lingoes, J. C "The Rationale of the Guttman-Lingoes Nonmetric Series. A Letter to Dr. Philip Runkel." *Multivariate Behavioral Research*, 3 (1968), 495–508.

35 Linsky, A. S. "Some Generalizations Concerning Primate Cities." *Annals of the Association of American Geographers*, Vol. 55, No. 3 (1965), 506–513.

36. Luce, R. D. *Individual Choice Behavior*. New York: Wiley, (1959).

37. Lycan, Richard. "An Analysis of Migrant's Distance and Preference Spaces in Canada, 1949-1968" (Paper read at Annual Meeting, Western Regional Science Association, San Diego, February 1970).

38. McGee, V. "The Multidimensional Analysis of 'Elastic' Distances," *The British Journal of Mathematical and Statistical Psychology*, 19 (1966), 181–196.

39 McGee, V. "The Multidimensional Scaling of N-Sets of Similarity Measures. A Nonmetric Individual Differences Approach," *Multivariate Behavioral Research*, 2 (1968), 233–248.

40 Mehta, S. K. "Some Demographic and Economic Correlates of Primate Cities," *Demography*, 1 (1964), 136–147.

41 Olsson, G. "Correspondence Rules in Geography." (mimeo), Ann Arbor, Mich., Department of Geography, University of Michigan, (1970).

42 Roskam, E. E. *Metric Analysis of Ordinal Data in Psychology* (Doctoral dissertation, University of Leiden), Voorschoten, The Netherlands, (1968).

43 Roskam, E. E. *A Comparison of Principles for Algorithm Construction in Nonmetric Scaling* Ann Arbor, Mich., Michigan Mathematical Psychology Program, No. 2, (1969).

44 Rushton, G "Analysis of Spatial Behavior by Revealed Space Preference," *Annals of the Association of American Geographers*, Vol. 59, No. 2 (1969), 391–400, (a).

45 Rushton, G "The Scaling of Locational Preferences," in K. Cox and R. G. Golledge (eds.) *Behavioral Problems in Geography*, A Symposium, Evanston, Ill., Northwestern University, Department of Geography, Studies in Geography, No. 17, (1969), 197–227, (b).

46 Rushton, G "Temporal Changes in Space-Preference Structures," *Proceedings of the Association of American Geographers*, Vol. 1 (1969), 129–132, (c).

47 Rushton, G "Preference and Choice in Different Environments," *Proceedings of the Association of American Geographers*, Vol. 3 (1971), 146–149.

48. Rushton, G. "Postulates of Central Place Theory and Properties of Central Place Systems," *Geographical Analysis*, III, 2 (April 1971), 140-156.

49 Schwind, Paul J "Spatial Preferences of Migrants for Regions. The Example of Maine," *Proceedings of the Association of American Geographers*, Vol. 3, (1971), 150–156.
50 Shepard, R N "The Analysis of Proximities Multidimensional Scaling with an Unknown Distance Function, I," *Psychometrika*, 27 (1962), 125–139; Part II, Psychometrika, 27 (1962), 219–246.
51 Shepard, R N. "Metric Structures in Ordinal Data," *Journal of Mathematical Psychology*, 3 (1966), 287–315.
52 Sherman, C R. *Nonmetric Multidimensional Scaling. The Role of the Minkowski Metric*, Chapel Hill, N.C.: Research Report No. 82, (1970), The L. L. Thurstone Psychometric Laboratory.
53. Spaeth, H. J. and S. B. Guthery "The Use and Utility of the Monotone Criterion in Multidimensional Scaling," *Multivariate Behavioral Research*, 4 (1969), 501–515.
54. Tobler, W R. "An Empirical Map Projection of the United States," Unpublished Manuscript, Ann Arbor, Mich.: Department of Geography, University of Michigan, (1967).
55 Tobler, W. R., H. W. Mielke, and T. R. Detwyler "Geobotanical Distance Between New Zealand and Neighboring Islands," *Bio Science*, (May 1970), 537–542.
56 Tobler, W R., and S. Wineberg "A Cappadocian Speculation," *Nature*, 231 (May 7, 1971), 39–41.
57 Torgerson, W. S. *Theory and Methods of Scaling*, New York: J. Wiley, (1958).
58. Whipple, T W. and L. A. Niedell "Black and White Perceptions of Competing Stores: An Application of Nonmetric Multidimensional Scaling," (mimeo), Department of Marketing, University of Buffalo, (1970).
59 Wilkinson, E. M. "Archaeological Seriation and the Travelling Salesman Problem," Paper to Bucharest Science Meeting, August 1971.
60. Young, F W "TORSCA-9: A Fortran IV Program for Nonmetric Multidimensional Scaling," *Behavioral Science*, 13 (1968), 343–344.
61 Young, F W. *A Fortran IV Program for Nonmetric Multidimensional Scaling*. Chapel Hill, N.C.: Report No. 56 (1968), The L. L. Thurstone Psychometric Laboratory.
62 Young, F W and M I. Appelbaum. *Nonmetric Multidimensional Scaling. The Relationship of Several Methods*. Chapel Hill, N.C.: Report No. 71, (1968), The L. L. Thurstone Psychometric Laboratory.
63. Young, F. W. and W. S. Torgerson. "TORSCA, A Fortran IV Program for Shepard-Kruskal Multidimensional Scaling Analysis," *Behavioral Science*, 12 (1967), 498.

Technical Paper Series

No. 1–Field Training in Geography, 1968
No. 2–Computer Assisted Instruction in Geography, 1969
No. 3–Evaluating Geography Courses. A Model with Illustrative Applications, 1970
No. 4–Living Maps of the Field Plotter, 1971
No. 5–Simulation of the Urban Environment, 1972
No. 6–Computerized Instruction in Undergraduate Geography, 1972
No. 7–The Interface as a Working Environment: A Purpose for Physical Geography, 1972
No. 8–LANDUSE: A Computer Program for Laboratory Use in Economic Geography Courses, 1972
No. 9–A Bibliography of Statistical Applications in Geography, 1972

Dr. John F. Lounsbury, Project Director, Arizona State University, Tempe, Arizona 85281, Dr. Salvatore J. Natoli, Associate Director, Association of American Geographers